DOCUMENT RESUME

ED 150 181 TM 006 904

TITLE	Georgia Educational Research Association: Presented Papers.
PUB DATE	28 Apr 77
NOTE	162p.; Papers presented at the Annual Meeting of the Georgia Educational Research Association (1st, Atlanta, Georgia, April 28, 1977)
EDRS PRICE	MF-$0.83 HC-$8.69 Plus Postage.
DESCRIPTORS	Affective Behavior; Answer Keys; *Conference Reports; Creativity; Criterion Referenced Tests; *Educational Research; *Elementary Education; Evaluation Methods; Higher Education; Individualized Instruction; Inservice Teacher Education; Measurement Techniques; Models; Personnel Evaluation; Privacy; Probability; Program Evaluation; Statistical Analysis; Student Evaluation of Teacher Performance; Teacher Attitudes; *Teacher Education; Testing Programs; Test Reliability; Test Validity
IDENTIFIERS	*Georgia; Georgia Criterion Referenced Test; Georgia Educational Research Association; School Survey (Coughlan and Cooke)

ABSTRACT

Ten papers presented at the first annual meeting of the Georgia Educational Research Association describe the following topics in educational research: (1) development and evaluation of the Elementary Curriculum Development Program, which involves individualized instruction in grades K-8; (2) reliability of the eighth grade reading subscales on the Georgia Criterion-Referenced Test; (3) comparison of anonymous and non-anonymous responses by public school teachers rating their satisfaction with their work environment and their school principals; (4) factor analysis and validity of the School Survey, also mentioned in the previous paper; (5) characteristics of beginning college students in elementary teacher education; (6) effects of an inservice creativity workshop on first grade teachers and their students; (7) a research and evaluation model for teacher education programs; (8) affective correlates of primary leval student behaviors; (9) stability of college students' ratings of instructors and courses; and (10) examination of the bilevel dimensionality of lower and higher mental processes in probability achievement. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED150181

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Georgia Educational Research Association

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Georgia Educational Research Association

Presented Papers

Annual Meeting
April 28, 1977

Georgia State University
Atlanta, Georgia

TM006 904

Does Individualized Instruction Make A Difference?

Dr. Nora Crawley
Dr. Lucille Jordan

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Nora Crawley
Lucille Jordan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Georgia Educational Research Association
April 28, 1977
Georgia State University
Atlanta, Georgia

DOES INDIVIDUALIZED INSTRUCTION MAKE A DIFFERENCE?

ELEMENTARY CURRICULUM DEVELOPMENT
DEPARTMENT OF PROGRAM DEVELOPMENT
DIVISION OF INSTRUCTIONAL PLANNING AND DEVELOPMENT
ATLANTA PUBLIC SCHOOLS

The Elementary Curriculum Development Program is a comprehensive effort to design and deliver a continuous progress, non-graded, individualized/personalized instructional program to Atlanta's elementary students. ECD is not a separate program--it is an umbrella--the total elementary curriculum and it was developed and implemented 1970-76.

ECD is a process curriculum model involving these concepts:

Components:

A. School Organization:

1. Leadership Team - comprised of one teacher from each instructional level (primary, intermediate, upper); principal, media resource teacher, a parent - area office and ECD person.

    a. meets at least once a week
    b. responsible for decision-making relative to curriculum and instruction

2. Instructional Teams - Primary, Intermediate, Upper teachers meet with instructional level peers at least once a week prior to leadership team meeting

3. Instructional Level

    a. Primary (K-3)
    b. Intermediate (4-5)
    c. Upper (6-8)

B. The curriculum is written based on application of the learning process in such a way that it is expected that pupils will develop autonomy in learning these Areas of Literacy:

1. Intellectual - 10 subject areas; ex. math, reading, science, P. E., etc.
2. Personal/Social - Affective development
3. Aesthetic - Drama, envioronmental and various art forms
4. Career

C. Guides (3):

Blue - Primary
Red - Intermediate
Green - Upper

Contains major concepts, general objective, and enabling objective in each content area and within the five literacies.

Checklists:

Sequence of skills for each content area on which teacher maintains each student's instructional progress.

D. New Report Card:

In lieu of "A", "B", ..., a condensation of checklist skills which offers parents more specific information relative to students' progress and development of skills.

E. Management System:

Total program uses a management plan, central, area and each ECD school has its own management system which is visibly displayed at the school. It includes, school activities coded to school objectives and persons bearing responsibility. Each participant in the program has a milestone chart geared to overall objectives and activities and he determines dates for expected accomplishment.

F. Staff Development:

Needs based staff development is planned reflecting competencies required for teachers to implement process model.
Each ECD school has one "contact" teacher identified for each content area. This person attends scheduled workshop and is then responsible to return to his school and duplicate the workshop and serve as a resource to his school staff in that area.

G. Dissemination:

A 20-minute slide/tape presentation is available in each area office which explains the total program in layman's language.

Back copies of ECD newsletters are available from ECD office.

H. Community Involvement:

Each school has a community Advisory Council made up of community, business and industrial representatives as well as parents and school personnel. Curriculum and instructional plans are tested with this group before decisions are made.

I. Competencies Project:

In conjunction with Guice Elementary School staff, staff members from the University of Georgia, and Atlanta Schools curriculum staff, teacher competencies and indicators of those competencies were identified which are necessary for teachers to implement such a program. Competencies lie in areas of diagnosing, prescribing, implementing, evaluating, humanizing, and professionalizing.

J. Research and evaluation designs were built in from the beginning and were implemented in 1974-75. Findings were that:

1. Pupils in new curriculum schools for three years scored higher on math, reading, and composite ITBS scores than control group pupils. The difference was significant at the .005 level.

2. Self concepts of upper elementary pupils were statistically significant at .005 level in favor of experimental pupils over control group pupils.

3. 1975 attendance of pupils in ten original ECD schools was much improved over the schools' 1974 attendance.

4. Elementary teachers in new curriculum schools increased amount of individualization of instruction from September to May at a rate significant at the .01 level.

K. Implementation:

ECD staff shifted activities, in light of system reorganization, to greater emphasis on planning and development and the area office resource teachers have assumed major responsibility for on-site implementation.

In 1972-73	Ten elementary schools were selected to pilot curriculum written to that point.
1973-74	20 schools piloted; of ten original schools each was paired with a school new to program to give leadership and to help with staff development.
1974-75	40 schools implementing following matching plan mentioned for 1973-74.
1975-76	80 schools implementing following matching plan mentioned for 1973-74.
1976-77	All 105 elementary schools implementing the curriculum.

Authors: William L. Curlette,
Georgia State University

Dorothy L. Brown
400 Chelsea Circle, N.E.
Atlanta, GA 30307

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
William L.
Curlette
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ESTIMATING THE RELIABILITY OF SELECTED READING SUBSCALES ON THE GEORGIA CRITERION-REFERENCED TEST

According to Glaser (1971), the concept of a criterion-referenced test (CRT) is "one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards." In contrast, norm referenced-measurement is often concerned with ranking individuals on some ability. To achieve reliable rankings in norm-referenced measurement, it is usually desirable to have relatively large variability in the observed score distribution. However, as Carver (1970) has pointed out, if one conceptualizes CRT reliability in terms of consistent classifications (above or below the specified performance level or cutoff score) for each examinee over repeated trials, then this consistency does not necessarily depend upon the variability of the test scores. Hence, reliability indices developed for norm-referenced tests may not be applicable to criterion-referenced tests.

As far as we can determine, evidence for the reliability of the eighth grade Georgia Criterion-Referenced Test (GCRT) is not generally available. Using a test-retest procedure, reliability indices are calculated for each of six different reading subscales (or objectives) for the eighth grade GCRT. In addition, the reliability of the subtest consisting of these six different subscales is also computed. For comparison purposes, the usual norm-referenced test-retest reliability is also reported.

The Georgia Criterion-Referenced Test

The Reading Test of the GCRT contains a total of 120 multiple choice items which are grouped into twenty subscales with from five to eight items in each subscale. To obtain information concerning the reliability of this test, objectives one through six were selected. Scores on these objectives or subscales and a subtest consisting of these six subscales are reported individually for each student. These objectives, the number of items in the subscales corresponding to each of these objectives, and the minimum number of correct items to pass each objective are presented in Table 1.

It should be noted that two of the six subscales require only two out of five items correct in order to pass the objectives measured by these subscales. Assuming that the response of a student who is guessing may be represented by the binomial distribution[1], the probability of achieving the objective by chance can be obtained. Since each item has four alternatives, there is a .36 probability that a student who has no knowledge of the answers to the questions will obtain a pass decision which is incorrect. If a student knew the answer to only one of the questions and guessed at the remaining questions for an objective, there is a .68 probability of obtaining a pass decision which is incorrect.

---

[1] Although this assumption is probably not realistic (Lord, 1975), it is used to obtain a rough approximation for the effect of guessing.

TABLE 1

Objectives and Cutoff Scores

Objective 1:	identifies the sounds of long and short vowels in words. At least 5 of 7 items must be answered correctly to pass.
Objective 2:	identifies sounds of two- and three-letter combinations of consonants in words. At least 6 of the 8 items must be answered correctly to pass.
Objective 3:	divides words into syllables. At least 2 of the 5 items must be answered correctly to pass.
Objective 4:	recognizes how the same word accented in different positions can have different meanings. At least 2 of the 5 items must be answered correctly to pass.
Objective 5:	identifies the sounds that vowels have when they appear before R, L, W, and U and distinguishes between the two different sounds of C and G. At least 5 of the 7 items must be answered correctly to pass.
Objective 6:	identifies the phonetic spellings of words as they might appear in a dictionary. At least 3 of the 5 items must be answered correctly to pass.

Data

The tests were administered in a DeKalb County School which has a student population in grades eight through twelve of about 1500 students. The school is located in basically a working-to-middle-class, non-transitional neighborhood, and has a student body which encompasses a wide socio-economic range. The percentage of students passing objectives one through six in 1977 is compared to the 1976 county and statewide percentages in Table 2. Of the 305 students who took the first test, retest data was obtained for 263 students.

Procedures

The reading sections of the GCRT were administered on March 16-18, 1977, in the eighth grade English classes during regularly scheduled class meetings. The retest on objectives one through six was administered two weeks later on April 1 during regular class meetings of the eighth grade English classes. In order to provide uniformity of testing conditions, a training in-service meeting was held prior to the first test administration for all teachers who were to administer the tests. Each class section had the same test administrator for both the test and the retest.

Memory of the test items and intervening events between the test and retest appeared to have negligible effect on the retest scores. First, the week after the administration of the GCRT all eighth grade English classes were involved in a combination movie/

TABLE 2

Percentage of Students Passing Objectives One Through Six on GCRT

	Objectives					
	1	2	3	4	5	6
1976 Statewide	62	56	90	80	68	60
1976 DeKalb County	83	72	95	88	84	76
1977 DeKalb County School in study (n=305)	88	79	98	96	89	80
1977 data from DeKalb County School in study--first administration (n=263)	91	81	99	96	91	81

literature unit on The Diary of Anne Frank. During the second week, approximately one-half of the test group continued in a literature sequence while the other half received instruction in subject and verb agreement. Second, the means and standard deviations of the subscale scores for the test-retest data were very similar as shown in Table 3. Third, fewer than three percent of the students outwardly questioned the fact that the same test items were used on the retest.

Test-Retest CRT Reliability

When cut-off scores are available, CRT reliability may be conceptualized as the ability of the test to classify students consistently as achievers or nonachievers on two test administrations (Carver, 1970; Subkoviak, 1976). Hence, one index to estimate the CRT reliability is simply the proportion of examinees classified in the same state (achievers/nonachievers) on both the test and retest.

The formula for the coefficient of agreement is the proportion of students who were nonachievers on the first test and nonachievers on the second test plus the proportion of students who were achievers on the first test and achievers on the second test. Using the notation in Table 4, the coefficient of agreement, symbolized by Pa, is given by $Pa = a + d$ where a and d are proportions. For example, the CRT reliability for subscale 6 is .125 plus .677 which equals .802.

The frequency distributions in Table 4 may be used to illustrate a statistical aspect of CRT reliability. It turns out that

TABLE 3

Means and Standard Deviations on Test and Retest

		Test	Retest
Objective 1:	$\bar{X}$	6.17	6.24
	SD	1.20	1.21
Objective 2:	$\bar{X}$	6.71	6.61
	SD	1.57	1.68
Objective 3:	$\bar{X}$	4.26	4.26
	SD	.95	.98
Objective 4:	$\bar{X}$	3.89	3.76
	SD	1.23	1.40
Objective 5:	$\bar{X}$	6.32	6.11
	SD	1.34	1.54
Objective 6:	$\bar{X}$	3.67	3.58
	SD	1.32	1.51
Subtest (Objectives 1-6):	$\bar{X}$	31.02	30.56
	SD	5.48	6.24

TABLE 4

Frequency Distribution & Contingency Table for Subscale 3

Subscale 3
1st Admin.

	Score	Frequency
Non-Achievers	0	1
	1	1
Achievers	2	16
	3	29
	4	79
	5	137
		263

Subscale 3
2nd Admin.

	Nonachievers	Achievers	Total
Subscale 3 1st Admin. Non-Achievers	0 0/263= 0.0 (a)	2 2/263=.008 (b)	2 2/263=.008 (a+b)
Achievers	5 5/263=.019 (c)	256 256/263=.973 (d)	261 261/263=.992 (c+d)
Total	5 5/263=.019 (a+c)	258 258/263=.980 (b+d)	263

TABLE 4 (Cont.)

Frequency Distribution & Contingency Table for Subscale 6

Subscale 6
1st Admin.

	Score	Frequency
Non-Achievers	0	3
	1	20
	2	28
Achievers	3	50
	4	70
	5	92
		263

Subscale 6
2nd Admin.

Subscale 6 1st Admin.	Nonachievers	Achievers	Total
Non-Achievers	33 33/263=.125 (a)	18 18/263=.068 (b)	51 51/263=.194 (a+b)
Achievers	34 34/263=.129 (c)	178 178/263=.677 (d)	212 212/263=.806 (c+d)
Total	67 67/263=.255 (a+c)	196 196/263=.745 (b+d)	263

the more the scores tend to group near the cutoff score, usually the greater the chance of a person being classified differently on the second test administration. There are more people near the cut-off score on subscale 6 than on subscale 3. Thus comparing the frequency distributions for subscales 3 and 6 for the first test administration, one would expect that subscale 3 would have more people classified the same on the second administration. In fact, this occurred as may be seen by coefficients of agreement of .973 and .802 for subscales 3 and 6, respectively.

The coefficient of agreement does not take into account the fact that by chance alone some proportion of the students would probably be observed in cells a and d. To alleviate this problem, Swaminathan, Hambleton, and Algina (1974) suggest the use of Cohen's (1960) kappa (k) as another index of CRT reliability. Kappa is defined as

$$k = \frac{Pa - Pc}{1 - Pc}$$

where Pa is the proportion of consistent decisions observed and Pc is the proportion of consistent decisions by chance. In Table 5 the proportion of consistent decision by chance is given by $Pc = (a + b)(a + c) + (c + d)(b + d)$. Kappa is interpreted as the percentage of consistent decisions above the consistent decisions due to chance or the net gain in classification accuracy attributable to using the test.

Using the reliability procedures previously described, the coefficient of agreement and kappa were calculated for the six subscales and the subtest. These results in addition to the norm-referenced reliability, are given in Table 5. The norm-referenced

TABLE 5

A Comparison of Criterion-Referenced and Norm-Referenced Reliability Indices

	Criterion-Referenced		Norm-Referenced
	Coefficient of Agreement	Kappa	Test-Retest Correlation
Objective 1:	.91	.43	.61
Objective 2:	.89	.66	.70
Objective 3:	.97	.04	.53
Objective 4:	.90	.18	.57
Objective 5:	.92	.60	.67
Objective 6:	.80	.43	.60
Subtest:	.95	.70	.81

reliability for the subscales was low; however, the standard error of measurement was usually less than one. This can be explained by the small variability of the observed scores caused, in part, by the large number of people answering all items correctly on each subscale.

Conclusions

In general, the six reading subscales that were investigated exhibited high reliability in terms of the coefficient of agreement. Kappa for the subtest (objectives 1-6) appears satisfactory since it indicates that the test is classifying students 70 percent better than chance. However, the subscale kappa's are not always satisfactory from a statistical viewpoint.

Based on the data in Table 2, it does not appear reasonable to generalize these results to all schools involved in the statewide testing. Nevertheless, the results do present reliability information on the Georgia Criterion-Referenced Test that are probably relevant to schools having approximately the same percentage of students passing the subscales as the school in this study.

Although the two test administration method of estimating CRT reliability is more general since it requires fewer assumptions than single administration CRT reliability procedures, the latter are more economical. In the future, the appropriateness of the single administration reliability indices (Subkoviak, 1976; Huynh, 1976) could be compared to the test-retest reliability indices for sample data from statewide administration of the GCRT. If the comparisons were favorable, then the accumulation of reliability information would be facilitated by using a single administration reliability procedure.

REFERENCES

Carver, R. P. Special problems in measuring change with psychometric devices. In *Evaluative Research: Strategies and Methods.* Washington: American Institutes for Research, 1970.

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 20, 37-46.

Glaser, R. A criterion-referenced test. In Popham (Ed.) *Criterion-referenced Measurement.* New Jersey: Educational Technology Publications, 1971.

Huynh, H. On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement*, 1976, 13, 253-264.

Lord, F. M. Formula scoring and number-right scoring. *Journal of Educational Measurement*, 1975, 12, 7-11.

Subkoviak, M. J. Estimating reliability from a single administration of a criterion-referenced test, A paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

Swaminathan H. Hambleton, R. K. and Algina, J. A bayesian decision-theoretic procedure for use with criterion-referenced tests. *Journal of Educational Measurement*, 1975, 12, 87-98.

The Effects of Teacher Anonymity on Teacher Attitudes Toward Characteristics of the School Environment and Principals' Competencies: "The Immunity from Persecution Phenomenon"

Chad D. Ellett and David T. McCoy

University of Georgia

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Chad D. Ellett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the annual meeting of the
Georgia Educational Research Association

Atlanta, Georgia

April, 1977

Educational researchers and psychologists, more often than not, request personal identity, background, and personal characteristics information from subjects administered paper and pencil instruments and various psychological tests. This information is usually considered desirable for sample documentation, for investigating hypotheses related to personal characteristics and demographic variables, and for fulfilling the researcher's needs to gather as much information as possible for "exploratory" investigations of one kind or another. It is questionable whether this traditional practice leads to more or less reliable and valid information supplied by survey participants, and even whether the practice under all conditions meets ethical guidelines recommended by various professional associations.

Though many studies of the influence of extraneous variables on test scores have appeared in the professional literature, few investigations have attempted to assess the effects of personal anonymity and non-anonymity on respondents' answers to questionnaires, rating scales, attitude inventories, and psychological tests. Since a large amount of research in education, sociology, and psychology is conducted with these kinds of instruments, identifying information of one kind or another is frequently requested. Given that such investigations are widely reported in the literature and subsequent significant findings represent current levels of knowledge about particular phenomena, the effects of anonymity on subjects' responses is a matter of legitimate concern. The paucity of research in the area has provided little understanding of anonymity and non-anonymity effects, and most studies have yielded somewhat conflicting results.

Olsen (1936), used the Woodworth-Mathews Personal Data Sheet to assess the effects of anonymity on student responses and concluded that the probability that more personal symptoms will be reported is higher under anonymous conditions. A study by Fischer (1946) using the Mooney Problem Check List with college students indicated that the mean number of problems reported by students did not significantly differ between anonymous and name conditions, but that the number of serious problems reported was greater under conditions of anonymity. A study by Feather (1973) compared male and female secondary students' responses to the Rokeach Value Survey, the Cornell Job-Description Index, and a rating of happiness with school under name and anonymous conditions. Results indicated that response anonymity had little effect on the ranking of values though boys ranked being "independent" higher in the name, than in the anonymous condition. In the name condition boys gave more favorable responses to classmates and the typical teacher and reported greater happiness with school, but girls in the name condition were less favorable toward schoolwork and reported less happiness with school. In a study by Fuller (1974), the Navy Personnel Survey was administered to officers and enlisted men in both name and anonymous conditions. General differences in response to attitude questions were low in relative frequency and small in magnitude. Pro-Navy statements were endorsed by a higher proportion of officers identifying their answer sheets, and negative statements were endorsed by a higher proportion of the officers not identifying their answer sheets. None of the name and anonymous group differences for enlisted men were large enough to be of practical significance. A study by Elinson and Haines (1950), however, indicated a significantly higher proportion of pro-military attitudes among Air Force and Army enlisted men under identified rather than anonymous conditions.

According to Anastasi (1968) and Guilford (1954), the tendency for subjects to "fake-good" on self report devices is motivated by the desire to make a favorable impression. While it is generally assumed that respondents are less likely to fake-good under anonymous response conditions, research using personality inventories and attitude questionnaires has revealed little difference between name and anonymous conditions. This finding is generally consistent for research studies in which the same individuals have been tested under both conditions (Corey, 1937; Gerberich and Mason, 1948), and when equivalent groups have been tested (Ash and Abrahamson, 1952; Hamel and Reif, 1952; Kosen, Kitchen, Kochen, & Stodolsky, 1970; Rosen, 1960).

The direction of the differences between anonymous and name conditions reported in the literature have generally been consistent with a faking-good or social desirability interpretation as suggested above. Becker and Bakal (1970), for example found that anonymous-identified conditions had less effect on subjects low in defensiveness than subjects high in defensiveness. Similarly, Rosen (1960) found that an identified group of students expressed more positive attitudes toward a specific course in which they were enrolled than an anonymous group, but there were no differences between the two conditions for attitudes toward reading. Benson (1941), in a survey of attitudes towards political candidates, found a greater number of "undecided" answers in response to personal interviews than to secret ballots.

Most studies investigating the effects of anonymity and identity on test responses and other measures have compared complete identification and anonymity conditions and the effects of these procedures on responses to one instrument. In addition, most studies use a test-retest procedure examining the amount of carry-over from one testing to another. One exception is a study by Hartnett and Seligsohn (1967) who investigated the effects of varying degrees of personal anonymity on students' responses to four different types of psychological questionnaires. Results were generally inconclusive though the authors stated that anonymity does seem to be more relevant when considering instruments of a more personal nature to the respondent.

The purpose of the present investigation was to compare public school teachers' responses under anonymous and complete information conditions to two instruments: 1) the School Survey, a multi-dimensional instrument to assess teacher attitudes towards characteristics of their working environment, and

2) the *Principal Performance Description Survey* - Teacher Form, a multi-dimensional instrument designed to assess teachers' ratings of high-priority competencies and performance indicators of public school principals. A unique aspect of the study unlike other investigations cited above, was that all subjects were requested to provide personal background information except name and were assured that personal anonymity would be maintained since the researchers were only interested in "group" responses. This procedure allowed for a comparison of a complete information teacher group with a totally anonymous teacher group...anonymous because they chose to delete the requested background information themselves. Since the response format of the two instruments used varied in terms of the "interpersonal" nature of the judgments being made by teacher respondents, and the purpose for which the data were being collected, it was hypothesized that teachers choosing to remain anonymous would be generally more negative than teachers providing complete background information. In addition, it was hypothesized that these effects would be larger for the instruments designed to measure teachers' ratings of their principal's competencies than for the *School Survey* (teacher attitude) instrument.

METHOD

Sample

The sample for the study was derived from a 1974-76 large scale field test of the Georgia Principal Assessment System (GPAS) designed to measure principal competencies developed through Project R.O.M.E. (Results Oriented Management in Education) at the University of Georgia. For the first analysis (comparison of complete information and anonymous teacher groups on the competency instrument), data were available from 160 teachers in 58 schools (44 elementary and 14 secondary). For the second analysis (comparison of teacher groups on the *School Survey*) data were available from 294 teachers in 74 elementary and 21 secondary schools. Individual schools from which the teachers were sampled represented a variety of urban-rural differences and socio-economic backgrounds.

Instrumentation

The Principal Performance Description Survey (PPDS) - Teacher Form instrument was developed through Project R.O.M.E. at the University of Georgia during 1974 - 1975. It contains 117 statements of principal behavior deemed essential and/or highly desirable for the effective operation of a school by 90% of a sample of 290 principals participating in a statewide verification survey. A complete description of the development and psychometric properties of the PPDS can be found in Payne, Ellett, Perkins, Klein, and Shellenberger, (1975), or Ellett (1976). The instrument requires that teachers make two judgments about their principal for each performance item: Frequency With Which Principal Performs Task from 1--Never to 5--Very Often, and Effectiveness With Which Principal Performs Task from 1--Ineffective to 5--Very Effective. Individual items on the PPDS - Teacher Form instrument are aggregated according to both Administrative Operations and Functional Areas of Administrative Responsibility composites to yield 11 subscales. The subscale structure, the number of items on each subscale, and sample items for the PPDS - Teacher Form instrument are presented in Appendix A.

The School Survey is a 118-item, 14 scale instrument developed by Coughlan and Cooke (1974), and designed as a general measure of teacher morale or work attitude. The response format for the School Survey requires teachers to make one of three judgments (Disagree, No Opinion, Agree) about each attitudinal statement. Data bearing on the psychometric properties of the instrument can be found in Ellett (1976) or Coughlan and Cooke (1974). The subscales, number of items on each, and sample items for the School Survey can be found in Appendix B.

Procedure

Teachers participating in the study were provided with the School Survey and the PPDS - Teacher Form instruments in an instrument package with a cover letter informing them of the general purpose of the larger field test of the GPAS. The letter encouraged teachers to complete all instrument items and to complete the requested background information sheet. Teachers were informed that the confidentiality of their responses would be maintained at all times, and that the background information was only to be used for subsequent research purposes. The instruments were completed on a take-home basis allowing from 3 to 7 days for completion. Teachers were instructed to return their completed instruments to the school librarian in a sealed envelope provided. No names were requested on the background information sheet. However, teachers were asked to provide the following: Name of School, Sex, Age, Years Teaching Experience, Grade Primarily Taught, Level of School Taught (Elementary, Middle, or Secondary), Major Subject Matter Area Taught.

Throughout the two years of the initial field test of the GPAS, data were collected for the two instruments from approximately 2,400 elementary and secondary teachers. For the PPDS - Teacher Form instrument analyses, 80 teachers were identified as providing no background information as requested, and were treated as an Anonymous teacher group. A comparison group (Complete Background Information) was selected at random within schools in the sample proportional to the distribution of teachers in the Anonymous group. For the School Survey instrument analyses, 147 teachers were identified as providing no background information as requested, and comprised the Anonymous group. Again, a comparison group (Complete Information) was selected at random within schools in the sample proportional to the distribution of teachers in the Anonymous group. This selection process resulted in an equal number of teachers in the Anonymous and Complete Information groups for subsequent statistical analyses of responses to the two instruments.

Statistical Analyses

In order to examine the research hypotheses relating to the Anonymous and

Complete Information teacher groups, responses to the PPDS - Teacher Form and School Survey instruments were subjected to a t test for independent means. Individual teachers rather than schools were the units of statistical analysis and t-test comparisons were made between groups for each of the subscales on the two instruments.

RESULTS

A summary of descriptive statistics and t-test values for the Anonymous and Complete Information teacher groups for frequency ratings of composites of the PPDS - Teacher Form instrument is presented in Table 1. Composite scores for the PPDS instrument were computed by summing ratings for the number of items comprising an individual composite for each individual teacher. Lower mean scores for any composite represent teachers' perceptions of principals as less frequently performing behaviors comprising each PPDS composite.

In comparing scores for the Anonymous (Group I) and Complete Information (Group II) teacher groups, several interesting findings emerged. First, the group of teachers choosing to remain anonymous rated their principals significantly lower on all possible subscales of the PPDS instrument than teachers providing complete background information. The mean differences were all highly significant, and in a direction predicted by the original hypotheses related to the two groups. Secondly, it appears that teachers in the Anonymous group, though more negative in their competency ratings, were slightly less cohesive in their judgments (larger standard deviations) than teachers in the Complete Information Group.

It will be recalled that the PPDS - Teacher Form instrument required teachers to make both frequency and effectiveness of performance judgments about their principal's performance using five-point rating scales. A summary of descriptive statistics and t-test values for the Anonymous and Complete Information teacher groups for effectiveness of performance ratings for each composite of the PPDS - Teacher Form instrument is presented in Table 2. As can be seen in the table, findings paralleling those for frequency ratings are evident. Teachers in the Anonymous group rated their principals as less effective than teachers in the Complete Information group. In addition, all mean comparisons were highly significant and in a direction predicted by the original hypotheses of the study. Considering the standard deviation of each composite as a rough index of cohesiveness in judgments, a trend similar to that for frequency ratings emerged, with Anonymous teachers being less cohesive in their judgments than teachers in Group II.

In comparing mean composite scores for Tables 1 and 2, it can be seen that means in Table 2 for effectiveness ratings are slightly lower than those for frequency of performance ratings in Table 1. This finding was expected since instrument instructions informed teachers to emit effectiveness ratings for any item rated a frequency of 1--Never.

A third comparison of Anonymous and Complete Information teacher groups was made using the responses of 294 teachers (n=147 in each group) to dimensions of the School Survey (teacher attitude) instrument. Table 3 presents descriptive statistics and t-test values for this comparison. As can be seen in the table, some differences between Anonymous and Complete Information groups were found, though these were not generally as frequent as those for the PPDS - Teacher Form (principal competency assessment) instrument. As regards specific dimensions of the School Survey, the Group I (Anonymous) teachers were significantly more negative in their attitudes towards Administrative Practices, School-Community Relations, Supervisory Relations, Colleague Relations, Voice in Educational Programs, and opportunities for professional Performance and Development than Group II (Complete Information) teachers. Again, for these significantly different School Survey dimensions, Group I teachers were slightly less cohesive in their judgments than Group II teachers, a finding similar in nature to those with the PPDS - Teacher Form instrument. Of the significant differences reported in Table 3, the largest were for the Administrative Practices and Voice in Educational Program dimensions.

DISCUSSION

The purpose of this investigation was to assess the effects of teacher anonymity on their attitudes towards characteristics of their working environment and their ratings of principals' competencies. It was hypothesized that teachers choosing to maintain anonymity would be significantly more negative in their responses to the two instruments utilized in the study than teachers providing complete background information as requested. In addition, it was predicted that differences between anonymous and complete information teacher groups would be greater (more frequent) for an instrument designed to measure teachers' ratings of principals' competencies than for a general teacher attitude inventory. Statistical analyses undertaken to test these hypotheses led to their confirmation. For both instruments on which anonymous and complete information teacher groups were compared, teachers choosing to maintain personal identity were significantly more negative than teachers providing requested background information.

Past research studies aimed at assessing the effects of anonymity on test, questionnaire, and survey responses have yielded conflicting results, and only served to show how anonymity effects seemingly interact with other variables such as subject characteristics, particular instrumentation used, etc. Past research has not yielded any general propositions related to the predicted impact of anonymity/non-anonymity variables on test responses, though the variable is seen as an important one. No studies in the research literature have utilized teacher responses to instruments like those used in the present investigation. Nor have they compared complete information groups of teachers (or other subjects) to anonymous groups of teachers..anonymous because they elected to maintain anonymity even though assured by the investigators that confidentiality would be maintained at all times.

The most frequent and outstanding results of the present study were those related to the PPDS - Teacher Form instrument designed as a measure of teachers' ratings of principals' competencies. For each of the possible PPDS - Teacher Form subscales, and for both frequency and effectiveness of performance ratings, teachers in the anonymous group rated their principals more negatively than teachers in the complete information group. When these results are viewed within the "faking-good" model proposed by Anastasi (1968) and others for understanding anonymity effects, some interesting propositions can be made.

First, it is generally assumed that subjects are less likely to "fake-good" (i.e., provide socially desirable responses) under anonymous response conditions. This raises the question as to whether teachers in the anonymous group in the present study were "faking-good" through their more negative evaluations of principals' behaviors. Were they more negative in their ratings because they saw this as socially desirable? The more simple explanation might be that teacher motivation to maintain personal anonymity is strongly associated with negative evaluations of others' (in this case the principal) behaviors. Being unconcerned about providing identifying information (i.e., having trust in and providing what researchers frequently request to document samples, etc.) is apparently strongly associated with more positive evaluations of others' performances.

What seems most important in understanding the results of this study is the motive to personally delete requested background information. Written comments made by individual teachers responding to the instrument package, and other observations as well, lead us to believe that a model of "evaluation apprehension" may be useful in explaining differences between the two teacher groups compared. Asking teachers to provide evaluative judgments about their principal probably leads to the desire on the part of some teachers to maintain complete anonymity, particularly if they intend to provide negative judgments. Teachers providing requested information, on the other hand, are probably not as apprehensive, and provide more positive judgments. The motive to remain or not remain anonymous may be instigated by the fear of being in a sense "found out" by those being rated (in this case the principal himself), a general observation that might be termed the "immunity from persecution phenomenon." The responses obtained from the teacher groups in this study suggest these relationships. The difficulty from a scientific viewpoint is attempting to determine which teacher group provided the most valid ratings of principals' performances. It may be that teachers in the complete identification group were "faking-good," contrary to the proposed relationship between test anonymity and socially desirable responding described by Anastasi (1968). The relationship between these variables may be reversed when the particular instrument is one measuring evaluations of others', rather than one's own characteristics or traits.

Some support is provided for the rationale above when viewing analyses related to the School Survey instrument presented in Table 3. This instrument is a more global measure of teachers' attitudes towards many dimensions of their work environment, and as such, is not as specifically "performance" or "evaluative-based" as the PPDS - Teacher Form instrument. Judgments are made one step removed from the interpersonal context in the school. Therefore, it was not surprising to find generally fewer and less significant differences between anonymous and complete information groups for this school attitude inventory. Considering group differences for the two instruments, it appears that anonymity and its maintenance becomes a more critical factor affecting teacher (and possibly other subjects") responses, as an instrument becomes more specifically evaluative in nature...particularly evaluative of others' competency and/or job-related performance.

TABLE 1

Summary of Descriptive Statistics and t-test Values for Anonymous (I) and Complete Information (II) Teacher Groups for Composites of the Principal Performance Description Survey (PPDS) - Teacher Form Instrument (Frequency Ratings)

PPDS Instrument Composites	Group I (n=80)		Group II (n=80)			
Administrative Operations	$\bar{X}$	S.D.	$\bar{X}$	S.D.	t	p
Data Collection (13)[1]	33.7	14.4	39.6	11.3	-2.87	.005
Planning (7)	17.9	7.8	21.9	6.1	-3.61	.000
Communicating (74)	198.0	79.5	244.6	59.3	-4.21	.000
Implementing (21)	58.5	23.7	69.9	16.2	-3.57	.000
Evaluating (2)	5.5	3.1	6.5	2.3	-2.20	.029
Functional Areas of Administrative Responsibility						
Curriculum and Instruction (35)	91.8	41.2	109.4	29.5	-3.11	.002
Staff Personnel (40)	109.1	41.2	134.2	30.1	-4.39	.000
Pupil Personnel (17)	47.5	19.4	58.7	14.2	-4.15	.000
School-Community Interface (1)	3.5	1.7	4.1	1.1	-3.09	.002
Fiscal Management (9)	21.6	10.2	27.2	9.6	-3.54	.001
System-wide Policies and Operations (15)	40.0	16.8	48.9	13.8	-3.66	.000

---

[1] Number of items in composite

TABLE 2

Summary of Descriptive Statistics and t-test Values for Anonymous (I) and Complete Information (II) Teacher Groups for Composites of the Principal Performance Description Survey (PPDS) - Teacher Form Instrument (Effectiveness Ratings)

PPDS Instrument Composites	Group I (n=80)		Group II (n=80)			
Administrative Operations	$\bar{X}$	S.D.	$\bar{X}$	S.D.	t	p
Data Collection (13)[1]	32.1	16.4	38.5	14.8	-2.65	.009
Planning (7)	16.0	9.1	21.1	8.1	-3.77	.000
Communicating (74)	188.2	85.1	238.7	76.6	-3.95	.000
Implementing (21)	55.2	24.4	67.6	20.9	-3.16	.001
Evaluating (2)	4.9	3.2	6.3	2.6	-3.16	.002
Functional Areas of Administrative Responsibility						
Curriculum and Instruction (35)	87.5	44.0	106.2	38.8	-2.86	.005
Staff Personnel (40)	102.7	43.3	130.9	39.2	-4.33	.000
Pupil Personnel (17)	45.4	21.2	57.5	18.2	-3.89	.000
School-Community Interface (1)	3.3	1.6	4.0	1.3	-2.85	.005
Fiscal Management (9)	19.4	11.5	26.5	11.4	-3.92	.000
System-wide Policies and Operations (15)	38.1	17.7	47.2	17.3	-3.29	.001

[1] Number of items in composites

TABLE 3

Summary of Descriptive Statistics and t-test Values for Anonymous (I) and Complete Information (II) Teacher Groups for Dimensions of the School Survey
(n=147 each group)

	Group I		Group II			
School Survey Subscales	$\bar{X}$	S.D.	$\bar{X}$	S.D.	t	p
Administrative Practices (9)[1]	17.78	5.46	20.14	4.91	-3.89	.000*
Professional Work Load (9)	20.64	5.18	21.52	4.64	-1.54	.124
Non-Professional Work Load (6)	13.42	3.61	14.00	3.07	-1.48	.140
Materials and Equipment (8)	17.00	5.33	15.91	4.70	1.87	.063
Buildings and Facilities (7)	15.37	4.73	15.55	4.66	-0.32	.747
Educational Effectiveness (10)	21.42	5.16	21.24	5.13	0.30	.768
Evaluation of Students (10)	21.16	5.97	21.71	5.41	-0.84	.402
Special Services (8)	15.42	4.76	15.97	4.54	-1.00	.317
School-Community Relations (7)	15.18	3.36	16.25	3.10	-2.82	.005*
Supervisory Relations (10)	22.61	5.82	24.01	5.63	-2.09	.038*
Colleague Relations (7)	15.05	5.32	16.39	4.31	-2.37	.018*
Voice in Educational Programs (8)	15.36	4.47	17.01	4.27	-3.24	.001*
Performance and Development (10)	21.01	5.50	22.31	5.40	-2.04	.042*
Financial Incentives (9)	14.33	5.24	15.42	5.27	-1.76	.079

[1] Number of items comprising subscale

*Means significantly different at $p<.05$ level

Appendix A

Composites and Sample Items for the Principal Performance Description Survey - Teacher Form Administered to Elementary and Secondary Teachers

Administrative Operations	Sample Item
1. Data Collection (13)[1]	Collects suggestions from staff as most effective methods for implementing board of education recommendations.
2. Planning (7)	Schedules staff meetings to discuss curriculum content areas where modification is needed.
3. Communicating (74)	Communicates to staff the importance of their participation in school policy decision making.
4. Implementing (21)	Appoints teachers to serve on curriculum committee.
5. Evaluating (2)	Evaluates instructional climate by observing in the classroom.
Functional Areas of Responsibility	
1. Curriculum and Instruction (39)	Discusses student performance criteria with teachers in individual conferences.
2. Staff Personnel (38)	Establishes a teacher committee to evaluate school's operation.
3. Pupil Personnel (22)	Develops individualized instruction programs based on student interests.
4. School-Community Interface (1)	Meets with teachers to discuss parental and community opinion about the school and its programs.
5. Fiscal Management (8)	Plans budget on projected needs data obtained from staff.
6. System-wide Policies and Operations (9)	Appoints staff committees to make recommendations for procedures to implement system goals in the school.

---

[1] Number of items in composite

Appendix B

Subscales and Sample Items of the School Survey Administered to Elementary and Secondary Teachers

Scales	Sample Item
1. Administrative Practices (9)[1]	"The administration seems willing to give careful consideration to our ideas and suggestions."
2. Professional Work Load (9)	"I am asked to spend too much time in meetings around here."
3. Non-Professional Work Load (6)	"I receive sufficient clerical assistance to do my job effectively."
4. Materials and Equipment (8)	"The instructional materials provided for me here are very satisfactory."
5. Buildings and Facilities (7)	"The layout of this school is inconvenient for the staff."
6. Educational Effectiveness (10)	"In my opinion, adequate educational standards are being upheld in this school."
7. Evaluation of Students (10)	"Students' absences are excessive in this school."
8. Special Services (8)	"Our library services for students are very satisfactory."
9. School-Community Relations (7)	"In general, I approve of school board policies."
10. Supervisory Relations (10)	"I seldom get the help I need in handling difficult discipline cases."
11. Colleague Relations (7)	"People in my school cooperate well."
12. Voice in Educational Program (8)	"I should have a greater voice in selecting students' textbooks and reference material."
13. Performance and Development (10)	"I think my work performance is judged fairly here."
14. Financial Incentives (9)	"I feel our salary system adequately rewards outstanding work."

---

[1] Number of items in subscale

References

Anastasi, A. Psychological Testing. (3rd ed.), New York: McMillan, 1968.

Ash, P., & Abrahamson, E. The effect of anonymity on attitude questionnaire response. Journal of Abnormal and Social Psychology, 1952, 47, 722 - 723.

Becher, G., & Badal, D.A. Subject anonymity and motivational distortion in self report data. Journal of Clinical Psychology, 1970, 26, 207 - 209.

Benson, L.E. Studies in secret ballot technique. Public Opinion Quarterly, 1941, 5, 79 - 82.

Corey, S.D. Signed versus unsigned attitude questionnaires. Journal of Educational Psychology, 1937, 28, 144 - 148.

Coughlan, R.J. & Cooke, R.A. Work Attitudes. In Walberg, H.J. (Ed.) Evaluating educational performance. Berkley, California: McCutchen Publishing Corporation, 1974, 295 - 317.

Elinson, J., & Haines, V.T. Role of anonymity in attitude surveys. American Psychologist, 1950, 5, 315.

Ellett, C.D. The Continued Refinement and Development of the Georgia Principal Assessment System and its Application to a Field-Based Training Program for Public School Principals: Assessment Design--Procedures--Instrumentation--Field Test Results. Vol 1, (Final Report), Georgia State Department of Education/Thomas County, Georgia, University of Georgia, 1976.

Feather, N.T. Effects of response anonymity on assessment of own and school value systems and satisfaction with school. British Journal of Educational Psychology, 1973, 43 (2), 140 - 150.

Fischer, R.P. Signed versus unsigned personal questionnaires. Journal of Applied Psychology, 1074, 30, 220 - 225.

Fuller, C. Effect of anonymity on return rate and response bias in a mail survey. Journal of Applied Psychology, 1974, 59 (3), 292 - 296.

Gerberich, J.B., & Mason, J.M. Signed versus unsigned questionnaires. Journal of Educational Research, 1948, 42, 122 - 126.

Guilford, J.P. Psychometric methods. (2nd ed.) New York: McGraw-Hill, 1954.

Hamel, L., & Reif, H.G. Should attitude questionnaires be signed? Personnel Psychology, 1952, 5, 87 - 92.

Hartnett, R.T., & Seligsohn, H.C. The effects of varying degrees of anonymity on responses to different types of psychological questionnaires. Journal of Educational Measurements, 4 (2), 1967.

Kosen, D., Kitchen, C., Kochen, M., & Stodolosky, D. Psychological testing by computer: Effect on response bias. Educational and Psychological Measurement, 1970, 30, 803 - 810.

Olsen, W.C. The waiver of signature in personal reports. Journal of Applied Psychology, 1936, 20, 442 - 450.

Payne, D., Ellett, C., Klein, A., Perkins, M., and Shellenberger, S. The Verification and Validation of Principal Competencies and Performance Indicators: Assessment Design--Procedures--Instrumentation--Field Test Results. Vol. 1, (Final Report), Georgia State Department of Education/Thomas County, Georgia, University of Georgia, 1976.

Rosen, N.A. Anonymity and attitude measurement. Public Opinion Quarterly, 1960, 24, 675 - 679.

The Structure of Teacher Attitudes Towards Dimensions of Their Working Environment: A Factor Analysis of the School Survey and Its Implications for Instrument Validity

Chad D. Ellett and J. Anthony Masters

University of Georgia

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Chad D. Ellett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the annual meeting of the

Georgia Educational Research Association

Atlanta, Georgia

April, 1977

During the past decade educational researchers have demonstrated a renewed interest in developing "high-inference" measures of characteristics of the school climate/learning environment. Part of this interest has seemingly been stimulated by developing models of environmental/ecological psychology and education and part by the failure of "low-inference" behavioral measures to reliably predict school outcomes of major research interest. Group perceptions and instrumentation to measure these, rather than behavioral checklists and rating scales, are again appearing in the research literature as valuable research tools. Much has been accomplished in the development of instrumentation for measuring students' perceptions of characteristics of the school climate/learning environment, but relatively little has been accomplished in the same area for teachers.

Several investigations (Walberg, 1968; Walberg & Anderson, 1969; Anderson, 1970) provide evidence that these student perceptions account for significantly more learning variance than more classically investigated characteristics such as I.Q. A recent study by Walberg, Signh, & Rasher (1977) indicates that student perceptions of characteristics of the school learning environment have cross-cultural incremental and predictive validity. While student studies of this type are relatively abundant, little has been accomplished within the measurement of teacher perceptions of school context characteristics and establishing their relationships to educational criteria of importance. One exception to the paucity of high-inference teacher attitude measures is the work of Coughlan and Cooke (1968) and their development of the School Survey, a 14 subscale teacher instrument.

The purpose of this paper is to report on a large scale factor analysis of the School Survey contributed by 958 teachers in the State of Georgia in 43 elementary schools. In addition, supplemental analyses attempted to demonstrate that a simpler structure of the School Survey would establish higher criterion-related validity coefficients than the original structure proposed by Coughlan and Cooke (1968). It was also of interest to assess the effects on criterion-related validity coefficients when school size and school socioeconomic status (SES) characteristics were controlled. It was also of interest to examine adjustments in the validity coefficients established when school size and socioeconomic characteristics were statistically controlled through partial correlation analyses.

It was hypothesized that criterion-related validity of the School Survey would be enhanced when items were realigned on subscales resulting from the factor analyses undertaken, and that these relationships would be relatively independent of school size and SES.

METHOD

Sample

The sample for the study consisted of 958 elementary school teachers in the State of Georgia derived from 43 elementary schools. The sample schools represented a variety of urban-rural differences and socioeconomic status characteristics. Data were collected through Project R.O.M.E. (Results Oriented Management in Education), a research and development effort at the University aimed at identifying and validating competencies of public school principals.

Instrumentation

The School Survey is a 118 item, paper-and-pencil type device that requires approximately 30 minutes to complete. The 118 items are distributed over 14 major subscales, and teachers are asked to respond to each by checking Agree, No Opinion, or Disagree. A listing of the subscales and sample items for the School Survey can be found in Appendix A. Internal consistency reliability coefficients (Cronbach's Alpha) within schools varied across subscales for the sample investigated from .58 to .85, with a median coefficient of .71 (Perkins, 1976).

Procedure

Teachers were administered the School Survey as part of a larger instrument package on a take-home basis, allowing from 5 to 7 days for completion. Teachers were asked to provide some background characteristics information with the instrument for sample documentation, but were assured that their responses were contributed anonymously, and that the researchers were only interested in collective responses, and had no interest in identifying individual teachers.

Criterion Measures

For the criterion-related validity portion of the study, criterion measures consisted of student scores on a standardized achievement test and school average daily attendance. For the elementary schools, achievement test scores were derived from the five subtests of the Iowa Tests of Basic Skills (ITBS), and the ITBS Battery Total score. The average daily attendance (ADA) figures, (expressed as a percentage of the maximum possible attendance), were derived for each school from the most recent twenty-day period coinciding with the time of actual data collection.

Statistical Analyses

Responses to the complete 118 item School Survey instrument for the sample of 958 teachers were first subjected to a principal components factor analysis with varimax rotation procedure to resolve for simple structure. Values of 1.00 were entered as initial communality estimates. Before the analysis all items on the instrument were coded in a manner consistent with the original scale names such that a maximum score of 3 (AGREE) for any item was associated with an increase in positive teacher perception of the dimension represented by the proposed School Survey scale label.

For the validation analyses, Pearson product-moment correlation coefficients were computed between the original 14 School Survey scales and the criterion variables of student achievement (ITBS Battery Total) and school ADA. In addition, correlations were computed between the 8-factor structure of the School Survey (emerging from the principal components) analysis and these same two criterion variables. To examine the independence of the validity coefficients established from school size and school socioeconomic characteristics, first and second-order partial correlation coefficients were computed between both the original 14-factor and 8-factor structures of the School Survey and the achievement and ADA criterion variables. The school size index was derived from the total number of students in each school and varied from 2,248 to 255 students, with a mean equal to 810. The SES index used was the percentage of students in each sample school receiving free school lunches. Specific criteria used in determining student qualification for free lunches were generally uniform across sample schools. Within the sample investigated, SES percentages varied from 81% to 2% with a mean equal to 28.3%.

RESULTS

Major results of the study are reported below in two sections. The first of these summarizes the findings from the principal components factor analysis to confirm the originally proposed structure of the School Survey instrument. The second section details results of the validation analyses for both the 14 and 8-factor structure of the School Survey, and the partial correlation analyses to assess the effects of controlling school size and SES characteristics on the criterion-related validity coefficients.

Factor Analysis

Results of the principal components analysis with orthogonal (varimax) rotation yielded only eight salient factors for the School Survey instrument (minimum eigen value 1.00) accounting for approximately 46% of the total test variance. Based on these results, a decision rule was structured for realigning and retaining items for an eight-factor structure of the instrument. Items failing to load at .40 or greater on any factor were eliminated. The decision was made to eliminate items loading on more than one factor, if the difference between their respective loadings was .10 or greater. In keeping with the assumptions of orthogonality, items loading in a bipolar manner were retained and placed on factors relative to their highest loading.

With the application of these selection criteria and decision rules, only 71 of the original 118 items remained to describe the eight most salient factors. Table 1 (following Appendix A) presents a summary of the significantly weighted School Survey items for each of the eight factors resulting from the principal components analysis as well as the direction and magnitude of their respective loadings. In addition, the original School Survey scale name with which the items were associated is included, as well as loadings and factor numbers for items loading on more than one factor. Eigen values for each factor are also provided at the top of each table page. The item classification presented in the table is based on the selection and placement criteria described above.

As can be seen in Table 1, the separate subscales of the School Survey as originally designated by Coughlan and Cooke (1968) are not independent in terms of their item clusterings. In addition, many of the original items are not loaded on one primary and clearly interpretable factor. As regards the original items, data reported here indicate that fully 47% (n=56) of the original 118 School Survey items show either no meaningful relationship to any of the eight factors derived, or are loaded on more than one factor. The School Survey instrument's simple structure suggests that it may more clearly measure only eight interpretable factors. No attempt has been made in this paper to "name" the theoretical factors on the basis of item content. However, it appears that items from a variety of the original scales have some intuitive content similarity.

Validation Analyses

Table 2 summarizes the zero, first, and second-order partial correlation coefficients established between the original structure (14-factor) of the School Survey instrument and the criterion variables of student achievement and school average daily attendance for the 43 elementary schools in the sample. In computing these coefficients school means were used as the units of analysis. The zero-order correlations represent the original correlations between the School Survey subscales and the two criterion variables. The first-order partial correlations were computed by controlling for school SES. The second-order partial correlation coefficients were computed by removing between school variance associated with both the SES and school size indices from the original relationship between the School Survey dimensions and the criterion variables.

As can be seen in Table 2, the original relationships between the 14 dimensions of the School Survey and the achievement criterion for 7 of the subscales were considerably reduced when between school variance associated with the school size and SES characteristics was removed. Though the magnitude of the Educational Effectiveness, Evaluation of Students, School Community Relations, Colleague Relations, and Financial Incentives correlations with achievement was reduced at the second order, they maintained statistical significance, supporting the criterion-related validity of these School Survey dimensions as originally defined by Coughlan and Cooke (1968).

For the ADA criterion, four of the original 14 School Survey factors maintained statistically significant relationships (Professional Work Load, r=.29; Non-Professional Work Load, r=.30; Educational Effectiveness, r=.46; and Evaluation of Students, r=.46) when the effects of SES and school size were statistically removed. Given both the criterion variables used in this study, it appears that the original School Survey scales of Educational Effectiveness and

Evaluation of Students as measures of teachers' attitudes towards their working environment have the greatest criterion-related validity.

In order to examine the validity of the 8-factor structure of the School Survey instrument resulting from the principal components factor analysis described above, zero, first, and second-order partial correlation coefficients were computed between the 8-factor instrument (with items aligned as presented in Table 1) and the two criterion variables of student achievement and school ADA. The results of these correlational analyses are presented in Table 3. As can be seen in the table, only two factors (factors 6 and 8) of the 8 factors identified maintained significant relationships to the achievement criterion when the effects of SES and school size were removed. This is evident when comparing findings for the zero and second-order relationships. When the 8-factor structure of the School Survey was correlated with the ADA criterion, only factor 5 maintained criterion-related validity at the second order. Apparently the originally established relationships between the 8 factors of the instrument (identified through the analyses reported here) and school ADA (4 of 8 statistically significant) have much in common with the SES and school size characteristics.

DISCUSSION

The working research hypothesis of this paper was that the criterion-related validity of the School Survey instrument would be enhanced when items on the separate subscales were clustered around the factor structure emerging from a simple principal components factor analysis. In addition, it was of interest to assess the relationships between the original 14-factor structure of the instrument and the criterion variables of student achievement and school ADA, when the effects of SES and school size were statistically controlled through partial correlation analyses.

The results of the analyses undertaken suggest that the original 14-factor structure of the instrument as proposed by Coughlan and Cooke (1968) maintains stronger global validity than the 8-factor structure suggested by the factor analysis of the instrument completed with a recent sample of responses from 958 elementary school teachers. This is a somewhat surprising finding, since one would predict that a simpler structure of the instrument (greater internal consistency of the items) should lead to greater validity of the type investigated. It may be that the intuitive structuring of the original School Survey scales on the basis of their content is a more valid procedure for development of instruments of this type than arriving at statistical structures through factor analytic techniques. One would assume that internally consistent items tending to load on clearly definable factors should enhance criterion-related validity. However, the results reported in this study tend to negate this assumption, at least for teacher attitude instruments like the School Survey. Some of the zero-order (original correlations) between the 8-factor structure of the School Survey and the criterion variables investigated were larger in magnitude than those established between these same variables and the 14-factor structure of the instrument. However, the statistical control of between school variance associated with the SES and school size indices apparently had a more deleterious effect on the validity coefficients for the 8-factor structure of the instrument than for the 14-factor structure. No simple explanation of this finding seems evident. However, what does seem evident is that the confirmatory factor analysis undertaken failed to corroborate Coughlan and Cooke's intuitive structure for the School Survey, but, at the same time, provided a structure for the instrument with less criterion-related validity.

It might be interesting in the future to undertake an investigation comparing the two instrument structures in terms of their incremental validity (stepwise regression analysis against the same two criterion variables used in the present study).

In comparing the two criterion variables used and their relationship to teacher attitudes as assessed by high inference measures such as the School Survey, it appears that the student achievement measure has more in common with teachers' school environment perceptions than school ADA. This is considered an important finding when considering the results of similar investigations designed to validate measures of student perceptions of characteristics of the school climate/learning environment.

A more wholistic model of environmental/ecological theory in educational research will probably need to integrate findings from studies attempting to validate both teacher and student perceptual measures. In this way, a working model of teacher and student perceptual factors "mediating" the school context might be forthcoming...one leading to better criterion-related predictions in educational settings.

References

Anderson, G.J. Effects of classroom social climate on individual learning. American Educational Research Journal, 1970, 7, 2, 135 - 152.

Coughlan, R., & Cooke, R. Work Attitudes. In H.J. Walberg (Ed.), Evaluating educational performance. Berkeley: McCutchan Publishing Co., 1974, 295 - 317.

Perkins, Mark L. Canonical correlation analysis of the relationship among school climate, teacher morale, and the educationally relevant performance of fourth-grade students. Unpublished doctoral dissertation, University of Georgia,1976.

Walberg, H.J. The social environment as a mediator of classroom learning. Journal of Educational Psychology, 1969, 60, 443 - 448.

Walberg, H.J., & Anderson, G.J. Classroom climate and individual learning. Journal of Educational Psychology, 1968, 59, 414 - 419.

Walberg, H.J., Singh, R., & Rasher, S.P. Predictive validity of student perception: a cross-cultural replication. American Educational Research Journal, 1977, 14, 1, 45 - 49.

Appendix A

Subscales and Sample Items of the School Survey Administered to Elementary and Secondary Teachers

Scales	Sample Item
1. Administrative Practices (9)[1]	"The administration seems willing to give careful consideration to our ideas and suggestions."
2. Professional Work Load (9)	"I am asked to spend too much time in meetings around here."
3. Non-Professional Work Load (6)	"I receive sufficient clerical assistance to do my job effectively."
4. Materials and Equipment (8)	"The instructional materials provided for me here are very satisfactory."
5. Buildings and Facilities (7)	"The layout of this school is inconvenient for the staff."
6. Educational Effectiveness (10)	"In my opinion, adequate educational standards are being upheld in this school."
7. Evaluation of Students (10)	"Students' absences are excessive in this school."
8. Special Services (8)	"Our library services for students are very satisfactory."
9. School-Community Relations (7)	"In general, I approve of school board policies."
10. Supervisory Relations (10)	"I seldom get the help I need in handling difficult discipline cases."
11. Colleague Relations (7)	"People in my school cooperate well."
12. Voice in Educational Program (8)	"I should have a greater voice in selecting students' textbooks and reference material."
13. Performance and Development (10)	"I think my work performance is judged fairly here."
14. Financial Incentives (9)	"I feel our salary system adequately rewards outstanding work."

---

[1] Number of items in subscale

TABLE 1

FACTOR 1 (Eigen value = 16.63)

Item		Scale Name	Loading	Additional Loading
13.	My principal backs me up in my dealings with parents.	Supervisory Relations	.62	
21.	I think my performance is evaluated fairly by those who are assigned to appraise my work.	Perf. and Dev.	.57	
28.	My principal seldom tries to get my ideas about things.	Supv. Relations	.74	
41.	I'm rarely told whether or not I'm doing good work.	Perf. & Dev.	.54	
44.	I fail to understand how my work performance is appraised and evaluated in this school.	Perf. & Dev.	.51	
46.	I have adequate opportunity to express my viewpoints about the philosophy and goals of this school.	Voice in Ed. Prog.	.57	.35 (2)
51.	My principal keeps me well informed about matters affecting my work.	Supv. Relations	.66	
61.	I seldom get the help I need in handling difficult discipline cases.	Supv. Relations	.60	
69.	Most of the time it's safe to say what you think around here.	Colleague Relations	.52	.32 (2)
73.	Administrative matters seem to get more attention in this school than the educational program.	Admin. Practices	.44	.31 (2)
78.	My principal fails to "go to bat" for us with his superiors.	Supv. Relations	.66	
83.	My principal seldom shows initiative in seeking ways to help us in our work.	Supv. Relations	.67	
89.	My principal is fair in his dealings with me.	Supv. Relations	.69	
106.	My principal seems to take suggestions for improvement as a personal criticism.	Supv. Relations	.65	
113.	My principal has an unrealistic view of what goes on in my work situation.	Supv. Relations	.64	

TABLE 1 (cont.)

FACTOR 1

Item	Scale Name	Loading	Additional Loading and (Factor Number)
118. I would definitely recommend this school to prospective teachers as a good place to work.	Perf. & Dev.	.52	.47 (6)

TABLE 1 (cont.)

FACTOR 2 (Eigen value = 3.85)

Item	Scale Name	Loading	Additional Loading and (Factor Number)
4. I have plenty of opportunity to express my ideas about salary matters in this district.	Financial Inc.	.45	
6. It seems to me that the school board should reconsider the amount of authority it has delegated to the superintendent.	School/Community	.62	
7. Insofar as they affect me, decisions made by the superintendent are fair and equitable.	Admin. Practices	.63	
16. From all I can gather, people who get promotions in this school district deserve them.	Perf. & Dev.	.42	
19. The school board seems more concerned about keeping costs down than about building an effective school program.	Admin. Practices	.41	.40 (3)
27. I think the school board does all it can to help build an effective educational program.	Admin. Practices	.58	.31 (3)
34. In general, I approve of the educational policies of the school board.	School/Community	.55	
36. It seems to me that the school board fails to concern itself with some really important educational matters.	Admin. Practices	.59	.30 (3)
50. The superintendent seems to be willing to give careful consideration to the ideas and suggestions of teachers in this school.	Admin. Practices	.70	
57. The superintendent seems to have an effective working relationship with the school board.	School/Community	.50	
62. The school board seems to recognize the professional character of our work in the schools.	Admin. Practices	.57	

TABLE 1 (cont.)

FACTOR 2

	Item	Scale Name	Loading	Additional Loading and (Factor Number)
68.	The superintendent seems to lack interest in the personal welfare of the staff of this school.	Admin. Practices	.64	
70.	In my opion, the school board seems to be divided on too many issues.	School/Community	.48	
87.	The superintendent usually tries to take action on the compaints of staff members in this school.	Voice in Ed. Prog.	.58	
99.	There seems to be too much friction between administrators in this district.	Admin. Practices	.51	
107.	We are seldom informed about what the top administrators in this district are thinking.	Voice in Ed. Prog.	.47	

TABLE 1 (cont.)

FACTOR 3 (Eigen value = 3.05)

	Item	Scale Name	Loading	Additional Loading and (Factor Number)
9.	Our library services for students are very satisfactory.	Special Serv.	.56	
18.	The quality of supplementary materials for student use in this school needs to be greatly improved.	Materials & Equip.	.67	
23.	The instructional materials provided for me in my work are very satisfactory.	Materials & Equip.	.68	
24.	A student in this school sometimes has to do without needed supplementary materials.	Materials & Equip.	.63	
26.	The school library and/or reference materials available to students are adequate to meet instructional needs.	Materials & Equip.	.67	
35.	It is easy and convenient in this school to get teaching aids and equipment to use in the classroom.	Materials & Equip.	.66	
45.	I have sufficient supplies for my work.	Materials & Equip.	.64	
101.	This school district lags behind other districts of comparable size and financial resources in introducing up-to-date materials and equipment.	Materials & Equip.	.42	

TABLE 1 (cont.)

FACTOR 4 (Eigen value = 2.23)

	Item	Scale Name	Loading	Additional Loading and (Factor Number)
2.	Too many teachers in this school seem to be more concerned with their own personal interests rather than with the overall welfare of the school.	Colleague Rel.	.68	
76.	People in this school cooperate well.	Colleague Rel.	.70	
84.	There are many cliques or groups in this school that create an unfriendly atmosphere.	Colleague Rel.	.63	
88.	The poor work performance of some people on this school staff makes it difficult for us to achieve adequate instructional goals.	Colleague Rel.	.58	
91.	A few of the people in this school think they run the place.	Colleague Rel.	.44	.31 (1)
114.	Teachers and other professional personnel in this school freely share ideas and materials.	Colleague Rel.	.49	

TABLE 1 (cont.)

FACTOR 5 (Eigen value = 2.01)

	Item	Scale Name	Loading	Additional Loading and (Factor Number)
33.	I would prefer a different work assignment (grade level or subject matter) from the one I have now.	Prof. Work Load	.49	
64.	The number of students I have to work with makes it difficult for me to do a good job.	Prof. Work Load	.52	
71.	In working with my students, I have adequate opportunity to allow for their individual differences.	Prof. Work Load	.49	
77.	The students I work with seem to need an unusual amount of discipline.	Educ. Effect.	.68	
81.	Generally speaking, I feel I could do far better work with students different from those usually assigned to me.	Prof. Work Load	.52	
95.	My professional work load is fair and reasonable.	Prof. Work Load	.48	
116.	My students show normal consideration, courtesy, and respect.	Eval. of Students	.64	

TABLE 1 (cont.)

FACTOR 6 (Eigen value = 1.68)

	Item	Scale Name	Loading	Additional Loading and (Factor Number)
1.	People in the community this school serves are "education oriented".	Educ. Effect.	.64	
39.	Almost all students in this school seem well prepared for advancement to the next higher grade level.	Educ. Effect.	.56	
40.	In general, the parents of the students in this school are interested in helping us educate their children.	Educ. Effect.	.61	
56.	Relations between the parents of students and the staff of this school need to be improved.	Educ. Effect.	.44	
103.	Even when you take into account difference in student ability, other schools in this locality seem to be ahead of this one in educational effectiveness.	Educ. Effect.	.42	
105.	I feel our school system is one big reason why people choose to live in this community.	School/Community	.65	
108.	I would rate this school system as one of the best for those who want to work in education.	Voice in Ed. Prog.	.50	.38 (2)
111.	In my opinion, adequate educational standards are being upheld in this school.	Educ. Effect.	.54	

TABLE 1 (cont.)

FACTOR 7 (Eigen value = 1.64)

	Item	Scale Name	Loading	Additional Loading and (Factor Number)
14.	Physical facilities for our personal use (lounge, washroom, etc.) need to be greatly improved.	Build. & Fac.	.47	
79.	The buildings and grounds where I work are kept as clean and attractive as possible.	Build. & Fac.	.48	
86.	Adequate facilities are available for my use during off-periods for grading papers, meeting with students and parents, and the like.	Build. & Fac.	.51	
90.	The general physical condition of my classrooms (lighting, temperature, ventilation, etc.) hampers me in doing a good job.	Build. & Fac.	.50	
97.	The classrooms, offices, and other work areas in this school need considerable improvement.	Build. & Fac.	.67	
102.	The layout of this school is inconvenient for the staff.	Build. & Fac.	.55	

TABLE 1 (cont.)

FACTOR 8 (Eigen value = 1.43)

	Item	Scale Name	Loading	Additional Loading and (Factor Number)
48.	Our system for reporting student progress to parents needs considerable improvement.	Eval. of Students	.81	
54.	The specialized programs in this school (music, art, drama, physical education, etc.) need to be greatly improved.	Special Serv.	.40	
72.	We lack satisfactory procedures in this school for evaluating student progress.	Eval. of Students	.60	
82.	Our practices for marking and reporting student progress are satisfactory.	Eval. of Students	.79	

TABLE 2

Summary of Zero, First, and Second Order Partial Correlation Coefficients Between Dimensions of the School Survey and Student Achievement and Average Daily Attendance (n=43 elementary schools)[1]

	ITBS Battery Total			Average Daily Attendance		
School Survey Factors	Zero Order	1st Order[2]	2nd Order[3]	Zero Order	1st Order	2nd Order
1. Administrative Practices	.48*	.28*	.24	.34*	.25*	.19
2. Professional Work Load	.49*	.20	.19	.38*	.28*	.29*
3. Non-Professional Work Load	.39*	.20	.18	.38*	.31*	.30*
4. Materials and Equipment	.34*	.09	.10	.16	.05	.08
5. Buildings and Facilities	.44*	.24	.27*	.25*	.16	.22
6. Educational Effectiveness	.77*	.52*	.49*	.54*	.49*	.46*
7. Evaluation of Students	.67*	.44*	.43*	.50*	.43*	.42*
8. Special Services	.58*	.24	.26*	.11	-.08	-.06

---

[1] School means used as the units of statistical analysis

[2] First order computed by controlling for school socioeconomic status

[3] Second order computed by controlling for school socioeconomic status and school size

*Correlation significantly different (p<.05) from zero

TABLE 2 (cont.)

School Survey Factors	ITBS Battery Total Zero Order	ITBS Battery Total 1st Order[2]	ITBS Battery Total 2nd Order[3]	Average Daily Attendance Zero Order	Average Daily Attendance 1st Order	Average Daily Attendance 2nd Order
9. School-Community Relations	.45*	.27*	.25*	.23	.14	.09
10. Supervisory Relations	.24	.22	.19	.27*	.24	.18
11. Colleague Relations	.61*	.41*	.37*	.31*	.20	.09
12. Voice in Educational Program	.31*	.12	.08	.33*	.26*	.21
13. Performance and Development	.37*	.20	.19	.20	.12	.09
14. Financial Incentives	.55*	.43*	.42*	.02	-.10	-.15

---

[1] School means used as the units of statistical analysis

[2] First order computed by controlling for school socioeconomic status

[3] Second order computed by controlling for school socioeconomic status and school size

[*] Correlation significantly different ($p<.05$) from zero

TABLE 3

Summary of Zero, First, and Second Order Partial Correlation Coefficients Between An Eight-Factor Structure of the *School Survey* and Student Achievement and Average Daily Attendance (n=43 Elementary Schools)[1]

	ITBS Battery Total			Average Daily Attendance		
School Survey Factors	Zero Order	1st Order[2]	2nd Order[3]	Zero Order	1st Order	2nd Order
1	.42*	.28*	.24	.23	.14	.05
2	.55*	.28*	.23	.37*	.26*	.17
3	.43*	.08	.09	.13	-.03	-.01
4	.53*	.30*	.25	.26*	.14	-.01
5	.59*	.18	.16	.45*	.34*	.33*
6	.87*	.66*	.67*	.40*	.27*	.25
7	.51*	.19	.23	.22	.08	.15
8	.58*	.33*	.36*	.06	-.13	-.10

---

[1] School means used as the units of statistical analysis

[2] First order computed by controlling for school socioeconomic status

[3] Second order computed by controlling for school socioeconomic status and school size

*Correlation significantly different ($p < .05$) from zero

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

DESCRIPTIVE DATA OF BEGINNING STUDENTS IN TEACHER EDUCATION

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Randy F.
Elmore

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

A Paper Presented at the
Georgia Educational Research Association

By

Randy F. Elmore
Chad D. Ellett
Julie A. Elfman
Kenneth W. Pool

College of Education
University of Georgia
Athens, Georgia

During the school year 1974-75, the Program Research and Evaluation Committee (PRECO) of the University of Georgia's Division of Elementary Education began collecting entry data from students initiating degree programs within the division (Early Childhood, Elementary, and Middle School Education). Instrumentation includes the University of Georgia Entering Student Information Inventory; California Test of Mental Maturity, Form S, Level 5; the Cattell Sixteen Personality Factor Questionnaire; Brown Personal Beliefs Inventory; Brown Teacher Practices Inventory; Rokeach Dogmatism Scale; Rokeach Value Survey, Form D; and the University of Georgia Semantic Differential Inventory.

A summary of these data for students entering the division from 1974-1976 and national norms are presented.

Descriptive Data on Beginning Students in Teacher Education

The University of Georgia Committee for Program Research and Evaluation (PRECO) has begun an initial investigation of the role that specified programmatic variables and personal student characteristics play in determining program effectiveness. Recognizing the need for a comprehensive assessment system to complement other aspects of the Georgia teacher education model, a preliminary evaluation plan has been structured by PRECO. The evaluation plan includes the following areas of investigation: (a) personal characteristics of preservice teachers; (b) components of programs in professional sequences; (c) settings in which field and campus experiences occur; (d) procedures for monitoring progress of preservice teachers; (e) procedures for program development; (f) performance of pupils taught by preservice teachers; and (g) long range effects of teacher education programs on graduates.

Baseline information has been collected on 296 students entering University of Georgia programs since the Fall of 1974 including: (a) Entering Student Information Inventory (biographic and demographic data), (b) Sixteen Personality Factor Questionnaire (16PF); (c) California Test of Mental Maturity; (d) Personal Beliefs Inventory; (e) Teacher Practices Inventory; (f) Dogmatism Scale (D Scale); (g) Value Survey; and (h) the University of Georgia Semantic-Differential Inventory (10 teacher education concepts).

PRECO efforts to date have been guided by the belief that an accurate appraisal of entering students is a crucial first step in building a comprehensive assessment system. If there is to be eventual success in improving teacher preparation programs, a more comprehensive program evaluation model needs to be designed and implemented. It is expected that continuous assessments of students as they progress through the training sequence, as well as "on-the-job" assessments of graduates will become integral parts of a more comprehensive program assessment and evaluation model.

Current Data

In the following discussion, a brief description of each instrument currently being utilized to collect information from students will be presented in addition to the data obtained.

University of Georgia Entering Student Information Inventory

A longitudinal investigation into the characteristics and effectiveness of teachers and teacher education programs necessitated the collection and availability of a complete set of entry descriptive data for each student. The Entering Student Information Inventory was designed to obtain the following information: (a) biographic and demographic information; (b) data regarding career-related experience and career aspirations; (c) expectations regarding teaching strengths and weaknesses; and (d) type of program in which the student is enrolled within the Division.

The Entering Student Information Inventory is a twenty-four item instrument. Most of the responses are provided by the students individually or in group sessions. The items include:

a. Name
b. Identification number
c. Sex
d. Race
e. Year of birth
f. Place of birth (state)
g. Marital status
h. Number of children
i. Self-judged socioeconomic status for him/herself
j. Date he/she entered UGA teacher education program
k. Area of teacher education (i.e., elementary, early childhood, or middle school)

l. Type of program in which enrolled (i.e., conventional, competency-based, or field-staged)
m. Name of program advisor
n. If a transfer student from another college, information concerning the transfer.
o. Experiences working with children
p. Self-judged condition of his/her own health
q. Attitude toward the teaching of particular school subjects
r. Career aspirations

Items of information obtained from other sources also included on the inventory are:

a. Total number of accumulated college credits at time of admission to the program
b. Cumulative college grade-point average at time of admission to the program
c. Scholastic-Aptitude-Test (scores from high school)
d. Cumulative high school grade-point average at graduation
e. Scores obtained on the Georgia Regents Examinations (state-wide examinations for all sophomores in state-supported colleges and universities in Georgia)

The following summary table (Table 1) obtained from the Entering Student Information Inventory is based upon a sample of 137 students entering the division during the 1974-75 school year, and for 106 students entering during the 1975-76 school year.

Table 1

Summary Table of Elementary Education Division Entering Student Entry Data

1974-75 n = 137 Response	Category	1975-76 n = 106 Response
95%	White Single Female	96%
54%/19%	Born in GA/Born in other state	54%/25%
Excellent	Health	Excellent
84% Middle Class	S.E.S.	80% Middle Class
52%	Transfer to UGA	48%
2.68	Entry GPA to College of Education	2.8
2.9	High School GPA	3.0
441/462	SAT Verbal/Math	433.5/457.5
70%/25%/5%	Elem. Educ./ECE/Middle School	61%/33%/5%
96%/50%/35%	Prior Work with Children (babysitting/church school/ teacher aide)	98%/76%/37%
40%	College Leadership	28%

California Test of Mental Maturity

As current educational literature indicates, agreement has not yet been reached concerning the extent to which measures of intellectual development relate to success in teaching. On the one hand, there are those who consider a "substantial I.Q." necessary for a student to successfully satisfy the demands of a conventional teacher preparation program. On the other hand, there are those who consider an "I.Q." derived from raw scores obtained on a standardized group test of mental maturity an unreliable predictor of success. This issue was of considerable concern to staff members of the division as they reviewed the plans of the investigating committee. Most staff members expressed doubt regarding the use of the I.Q. as a predictor of professional success. Due to these conflicting notions and a dearth of evidence, it was decided to further investigate this issue and that the present longitudinal study would be an appropriate one in which to obtain the necessary data.

The California Short-Form Test of Mental Maturity was selected. The form of the test used was 1963 S-Form, Level 5, consisting of 120 multiple-choice items. This form is designed for group administration and is suitable for college-age persons.

The test purports to assess four factors by means of seven subtests. The first factor, logical reasoning, is measured by opposites, similarities, and analogy-type problems in picture form; and the second factor, numerical reasoning, by test of numerical values and by number problems. The third factor, verbal concepts, is assessed by verbal comprehension items; and the fourth factor, memory, by a test of delayed recall. The test is also subdivided to measure language as contrasted with nonlanguage skills. Tests concerned with the nonlanguage factor focused on opposites, similarities, analogies and numerical values and tests concerned with number problems, verbal comprehension and delayed recall yield the language factor score.

Table 2 compares the scores obtained by students entering the division during the 1974-75 and 1975-76 school years with the national norms for the adult population on the California Test of Mental Maturity (1963 S-Form, Level 5). In addition, I.Q. equivalents are well above the midpoint for the factors of language (113 and 114) and nonlanguage (111 and 114) and for the combined indicator (111 and 117).

In interpreting these data it must be realized that the subjects were entering students (i.e., transfer students or recently declared elementary division majors) rather than graduating students or established seniors. It is thus likely that many of these subjects will not complete the programs they have entered. Records obtained to date reveal that a number of students

Table 2

Comparison of Scores Obtained by Teacher Education Students (entering programs at the University of Georgia for the preparation of teachers of children) and National Norms for the Adult Population on the CTMM (1963 S-Form, Level 5)

Variable	Mean Raw Score		Range		S.D.		Percentile of adult Population		I.Q. Equiv.	
	1974-75	1975-76	1974-75	1975-76	1974-75	1975-76	1974-75	1975-76	1974-75	1976-77
Logical Reasoning	28.70	27.419	25.00	24.00	4.84	4.769	75	66		
Numerical Reasoning	15.60	15.210	19.00	17.00	3.73	3.754	73	69		
Verbal Concepts	15.85	15.762	20.00	20.00	3.91	3.754	82	76		
Memory	18.82	17.810	17.00	17.00	3.70	3.703	82	69		
Language	38.41	37.057	36.00	32.00	7.16	6.918	83	79	114	113
Non-Language	40.49	39.124	30.00	44.00	6.16	6.614	82	76	114	111
Total	78.90	73.371	60.00	53.00	11.60	10.728	86	76	117	111

in the lower quartile of the I.Q. distribution have withdrawn from programs in which they were enrolled. The observed range of 60 points on the CTMM represents a distribution from 79 to 139.

Sixteen Personality Factor Questionnaire (16 PF)

According to Cattell, et al., (1970), the 16 PF "is designed to make available, in a practicable testing time, information about an individual's standing on the majority of primary personality factors...on the human personality sphere. . ."

A multiple-choice instrument with 187 items, the 16PF can be administered to groups as well as individuals. It is intended for young adults and requires about 45 minutes to complete. (Scoring sorts for two sets of eight factors and the score for each factor is a standard summation.)

Figure 1 lists the 16 personality factors for which data are provided. The letters used as indicators for the 16 factors are the same as those used on the test and in the handbook (Cattell, et al., 1970). The names given in the table are the technical referents; those in parenthesis are less technical near-equivalents used in the handbook, where a more extensive description of each of the 16 characteristics is provided. It must also be mentioned that six to nine broader second-stratum factors have been derived from and accompany the 16 personality characteristics. Although these characteristics are not considered in this paper, further study may suggest the need for investigation.

Table 3 shows the mean scores of students entering the Division of Elementary Education from 1974-76. Comparisons are made between scores for 1974-75, and 1975-76, cumulative totals for both years, and scores for elementary teachers as reported by Cattell (1970, p. 185). Examination

Figure 1

Primary Source Traits Identified by the *16 PF*

Factor Referent	Characteristic for Low Score	Characteristic for High Score
A	Sizothymia (Reserved, detached)	Affectothymia (Warmhearted, outgoing)
B	Low Intelligence (Dull)	High Intelligence (Bright)
C	Emotional Instability (Affected by feelings)	Higher Ego Strength (Emotionally stable)
E	Submissiveness (Humble)	Dominance (Assertive)
F	Desurgency (Sober, taciturn)	Surgency Enthusiastic (Happy-go-lucky)
G	Low Superego Strength (Expedient)	Superego Strength (Conscientious)
H	Threctia (Shy, timid)	Parmia (Adventurous, socially bold)
I	Harria (Tough-minded)	Premisa (Tender-minded)
L	Alaxia (Trusting)	Protension(Suspecting)
M	Praxernia (Practical)	Autia (Imaginative)
N	Naivete (Forthright, unpretentious)	Shrewdness (Astute, worldly)
O	Untroubled Adequacy (Placid, self-assured)	Guilt Proneness (Apprehensive, self-reproaching)
Q_1	Conservatism of Temperament (Respecting established rules)	Radicalism (Experimenting, liberal)
Q_2	Group Adherence (Group dependency ("a joiner")	Self-sufficiency (Resourceful, prefers own decisions)
Q_3	Low Self-sentiment Integration (Uncontrolled)	High Strength of Self-sentiment (Controlled, exacting willpower)
Q_4	Low Ergic Tension (Relaxed, tranquil)	High Ergic Tension (Tense, frustrated)

reveals that scores for all groups fall within a "normal" range. The differences between 1974-75 and 1975-76 scores for University of Georgia students were small. On fourteen items the differences in mean factor scores ranged from .0 to .2. On two factors, H and I, the differences were .4, only one half a point. The consistency of the scores over the two years seems to support the reliability of the 16 PF.

In 1975, two notable differences were observed between University of Georgia students and elementary teachers: (a) on ten of the characteristics the entering students rate above the classroom teachers, and (b) on three of the characteristics, the discrepancy between mean scores for the student group and mean scores for the teacher group is approximately one or more stens. The three characteristics showing this difference were: (a) affected by feelings vs. emotionally stable, (b) uncontrolled vs. controlled, and (c) relaxed vs. tense.

In 1976, two differences were again observed between the scores of Georgia students and elementary teachers: (a) on ten of the factors entering students rate above the classroom teachers, and (b) on one of the characteristics, the discrepancy between the mean scores of entering students and the mean score of classroom teachers is more than one sten. The characteristic showing this difference was group adherence vs. self-sufficiency.

Personal Beliefs Inventory (PBI)

The PBI consists of 40 general philosophical statements taken from the writings of John Dewey. Twenty of these statements express ideas compatible with Dewey's philosophy, the other twenty express ideas incompatible with his philosophy (the compatibility of each statement being determined by a panel of six qualified judges).

Although some studies utilizing the PBI have separated the 40 items into two randomly arranged forms of twenty items each, it was decided to use all 40 statements for each administration of the instrument. In order to avoid response set or the possibility of group mean scores being affected by the order of item presentation, Brown has recommended that two forms of the test be prepared. Forms A and B were utilized, each containing all forty items in differing orders.

A total of 102 entering University of Georgia students responded to the instrument during 1974-75. The mean score was 131.4, with a standard deviation of 13.4. In 1975-76, 106 entering students achieved a mean of 132.4; the standard deviation was 14.5. Table 4 compares the responses of these entering students with the responses by a group of 154 undergraduate and graduate education students at the University of Wisconsin, Madison (Brown, 1968). As indicated, the means of the Georgia groups are below the mean of the Wisconsin group. It should be noted that all of the students in the Wisconsin group were in professional education programs, with some having teaching experience; those in the Georgia group were just entering professional education programs. On the basis of such limited data, further comment regarding scores appears to be unwarranted. It is anticipated that more meaningful comparisons will become possible as entering students complete their preparatory programs and gain teaching positions.

Table 3

Mean Scores on 16 PF Comparing Students
Entering Division of Elementary Education
and Elementary Teachers (Cattell et al., 1970)

	Entering Students[a]				Elementary Teachers[b]	
Factor	Mean		SD		Mean	SD
	74-75	75-76	74-75	75-76		
A	5.6	5.5	1.8	1.7	5.6	1.8
B	6.0	6.1	2.1	1.9	6.5	1.8
C	6.0	5.8	1.7	1.9	5.0	1.7
E	6.3	6.3	2.0	2.0	5.5	1.9
F	6.2	6.4	2.0	2.0	5.5	2.0
G	6.2	6.1	1.7	1.8	5.7	1.9
H	5.9	6.2	2.2	2.2	5.6	1.9
I	6.8	6.4	2.1	2.2	5.9	1.9
L	5.6	5.3	1.8	2.0	5.0	1.9
M	5.1	4.8	2.2	2.0	5.4	1.9
N	5.3	5.3	2.1	2.0	5.8	1.9
O	5.7	5.9	1.9	2.0	5.1	1.9
Q1	4.5	4.7	1.9	2.2	5.5	1.9
Q2	4.5	4.4	2.2	2.4	5.7	1.9
Q3	6.5	6.3	1.9	2.2	6.0	1.8
Q4	5.6	5.7	1.9	2.1	5.3	1.8

[a1] *n* = 103

[a2] *n* = 106

[b] *n* = 1280

Table 4

Results of *Personal Beliefs Inventory* Comparing Scores of Students Entering Division of Elementary Education at the University of Georgia with Scores of Education Students at the University of Wisconsin, Madison

Group	N	Mean	S.D.
University of Wisconsin Students	154	144.0	18.6
University of Georgia Students(1974-75)	102	131.4	13.4
University of Georgia Students(1975-76)	106	132.4	14.5

Teacher Practices Inventory (TPI)

To determine the general nature of the educational beliefs of entering students, another part of the *Experimental Scale* (Brown, 1962), entitled the *Teacher Practices Inventory* (TIP), was used in the current investigation.

The *TPI* consists of 40 statements concerning teaching practices suggested by various categories constituting a theoretical framework for describing John Dewey's educational philosophy. As with the *PBI*, all 40 items were used each time the *TPI* was administered; two forms of the *TPI* were provided, differing from each other only in the distribution of items.

Consistent with the PBI procedures, respondents to the *TPI* are requested to indicate on a six-point scale the extent of agreement each statement has with their own beliefs; scoring is congruent with the *PBI*, high scores indicating relative agreement with Dewey and low scores indicating a measure of disagreement.

In 1974-75, a total of 102 entering University of Georgia students completed the instrument with 106 students responding in 1975-76. The mean score in 1974-75 was 160.5, with a standard deviation of 21.4; and in 1975-76 the mean was 160.0, with a standard deviation of 20.9. Table 5 compares the responses obtained from this group of entering students with the responses made by a group of 154 education students at the University of Wisconsin, Madison (Brown, 1968). As indicated in the table, the mean of the Wisconsin group exceeds the mean of the Georgia group by 16 points. Again it is anticipated that more meaningful comparisons will become available at a later point in the longitudinal study.

Table 5

Results of Teacher Practices Inventory Comparing Scores of Students Entering Division of Elementary Education at the University of Georgia with Scores of Education Students at the University of Wisconsin, Madison

Group	N	Mean	S.D.
University of Wisconsin Students	154	176.7	20.4
University of Georgia Students(1974-75)	102	160.5	21.4
University of Georgia Students(1975-76)	106	160.0	20.9

Dogmatism Scale (D Scale)

The previously mentioned PBI and TPI both were selected in an attempt to measure the content of the basic philosophical and educational beliefs of entering students. However, the investigating committee also believed it necessary to attempt to measure the structure of these beliefs. A review of previous investigations relative to teacher education led to the selection of the Dogmatism Scale (D Scale) for this purpose.

The *D Scale*, developed by Milton Rokeach, purports to provide measures of the degree of flexibility characterizing a respondent's belief system. Studies using the *D Scale* indicate that it measures authoritarianism, i.e., the rigidity of a respondent's beliefs concerning people and ideas, independent of a particular ideology. When *D Scale* scores have been listed along with *PBI* and *TPI* scores, "...findings would indicate that the experimentalism inventories measure something that Rokeach's *Dogmatism Scale* does not, and *vice versa*" (Brown, 1968). Additionally, since the *D Scale* has been used extensively with college students, findings from the present investigation may be compared with similar studies conducted at other institutions.

The *D Scale* contains 40 statements and is designed to be administered to groups as well as to individuals. Respondents are asked to indicate, on a six-point scale, how much they agree or disagree with each statement. The score earned by a respondent is the summation of responses to all items; a higher score indicates a higher level of dogmatism, more closedness and greater rigidity, with a lower score indicating a lower level of dogmatism, more openness, and greater flexibility, in the structure of a respondent's belief system.

The mean score for the total of 102 respondents in 1974-75 was 136.0, with a standard deviation of 18.2. In 1975-76, the mean score for 106 respondents was 143.4 with a standard deviation of 20.3. Table 6 compares the responses of these groups with the responses of 154 undergraduate and graduate education students at the University of Wisconsin, Madison (Brown, 1968). As indicated in the table, a difference of six and thirteen points exists between the mean scores for the two groups for the two years. It is anticipated that the use of the D Scale with these groups of students, over a period of years, will provide useful information when combined with other data.

Table 6

Results of D Scale Comparing Scores
of Students Entering Division of Elementary
Education at the University of Georgia with Scores
of Education Students at the University of Wisconsin, Madison

Group	N	Mean	S.D.
University of Wisconsin Students	154	130.4	27.0
University of Georgia Students(1974-75)	102	136.0	18.2
University of Georgia Students(1975-76)	106	143.4	20.27

Value Survey

The Value Survey, Form D, is a projective test, simple in design and economical to administer to groups or individuals. It consists of two lists of 18 alphabetically arranged value referents. One list contains terminal values which are related to life goals both social and personal in orientation (e.g., "salvation," "family security," and "a world at peace") while the other list contains instrumental values which are related to life styles involving competence and morality (e.g., "honest," "logical," and "clean").

Respondents are instructed to arrange the values in each list in order of personal importance. The procedure assumes an importance exists in the relative ordering of values rather than in the presence or absence of any particular value. It does not assume that the two lists contain values which would be most important to every individual. The values in both lists would generally be considered socially desirable.

The instrument does not yield a conventional score. The rank-order procedure is ordinal, with each item capable of independence of meaning. Median scores may be derived to describe group rankings.

Median scores for the 18 terminal values of the Value Survey are shown in Table 7. Scores for students entering the division during 1974-75 and 1975-76 are compared with scores from one sample of only female subjects reported by the National Opinion Research Center and from another sample of both male and female subjects having some college education. The rank order of values for each group is given in parentheses.

Findings in Table 7 reveal certain similarities among the rankings of terminal values for all groups. However, certain differences also should be noted. Shown below in rank order are the six most important terminal values as named by each of the three groups:

			Entering College	
Rank	Female	Some College	1974-75	1975-76
1	A World at Peace	Family Security	Salvation	National Security
2	Family Security	A World at Peace	Family Security	Pleasure
3	Freedom	Freedom	Self-Respect	Mature Love
4	Salvation	Wisdom	Inner Harmony	Freedom
5	Happiness	Self-Respect	True Friendship	Salvation
6	Self Respect	A Sense of Accomplishment	Wisdom	Social Recognition

Table 7

Median Scores of Terminal Values on Rokeach Value Survey for Students Entering the Division Compared with Male and Female Subjects in a National Sample with Some College Experience and with Females in a National Sample*

	Terminal Values	Students Entering Division		National Sample	
		1974-1975	1975-1976	Female	Some College
1.	Comfortable Life	12.42 (13)	13.55 (17)	10.01 (13)	11.13 (13)
2.	Exciting Life	14.16 (15)	9.86 (12)	15.75 (18)	15.19 (18)
3.	Sense of Accomplishment	9.13 (10)	10.97 (14)	9.40 (10)	7.50 (6)
4.	A World at Peace	10.50 (11)	9.20 (10)	3.00 (1)	4.19 (2)
5.	A World of Beauty	12.97 (14)	9.68 (11)	13.51 (15)	13.63 (15)
6.	Equality	10.61 (12)	9.86 (12)	8.29 (8)	8.50 (8)
7.	Family Security	5.23 (2)	8.50 (8)	3.78 (2)	3.55 (1)
8.	Freedom	7.63 (9)	7.93 (4)	6.05 (3)	5.38 (3)
9.	Happiness	6.50 (7)	8.83 (9)	7.34 (5)	7.90 (7)
10.	Inner Harmony	5.57 (4)	8.39 (7)	9.83 (12)	9.50 (9)
11.	Mature Love	7.30 (8)	7.68 (3)	12.32 (14)	12.10 (14)
12.	National Security	15.00 (17)	6.90 (1)	9.81 (11)	10.10 (10)
13.	Pleasure	14.44 (16)	7.50 (2)	14.97 (16)	14.82 (16)
14.	Salvation	4.00 (1)	8.10 (5)	7.33 (4)	10.15 (11)
15.	Self Respect	5.44 (3)	10.67 (13)	7.40 (6)	6.88 (5)

Table 7

(Continued)

	Terminal Values	Students Entering Division 1974-1975	1975-1976	National Sample Female	Some College
16.	Social Recognition	15.17 (18)	8.30 (6)	15.01 (17)	15.09 (17)
17.	True Friendship	6.23 (5)	11.56 (15)	9.06 (9)	10.50 (12)
18.	Wisdom	6.23 (5)	11.78 (16)	7.71 (7)	6.04 (4)

Numbers in Parentheses Represent Rank-Order of Values for Each Group.

*Rokeach, M. The Nature of Human Values, New York: The Free Press, 1973, pp. 380-391.

"Family Security" and "Self-Respect" are among the most important values for females, college students, and entering students (1974-75). "Freedom" and "Salvation," rated highly by entering students (1975-76) are also among the highest values of three of the four groups. Only "Salvation" was in the highest values of both years of entering students.

The six terminal values considered to be least important by the two groups of entering students (1974-75 and 1975-76), female, and college students are listed below:

			Entering College	
Rank	Female	Some College	1974-75	1975-76
18	An Exciting Life	An Exciting Life	Social Recognition	Comfortable Life
17	Social Recognition	Social Recognition	National Security	Wisdom
16	Pleasure	Pleasure	Pleasure	True Friendship
15	A World of Beauty	A World of Beauty	An Exciting Life	Sense of Accomplishment
14	Mature Love	Mature Love	A World of Beauty	Self-Respect
13	A Comfortable Life	A Comfortable Life	A Comfortable Life	Exciting Life

During 1974-75, five of the six least important values were the same for entering students, females, and college students. The exception being entering students included "National Security" among the least important values, while both groups in the national sample did not. In 1975-76 two values, "Comfortable Life" and "Exciting Life" were ranked as least important by entering students and were also among the least important for all groups. However, the entering students (1975-76) rated "Wisdom," "True Friendship," and "Sense of Accomplishment" as least important which was different from the other three groups.

Median scores for the 18 instrumental values of the Value Survey are shown in Table 8. Scores for students entering the division during 1974-75 and 1975-76 are compared with scores for the same two groups -- that is, female subjects and both male and female subjects having some college education -- reported by the National Opinion Research Center. The rank ordering of values for each group is given in parentheses.

The six most important instrumental values as named by each of the groups include more similarities than differences. These six values, in rank order, are:

			Entering Students	
Rank	Female	Some College	1974-75	1975-76
1	Honest	Honest	Honest	Helpful
2	Forgiving	Responsible	Loving	Loving
3	Responsible	Courageous	Responsible	Honest
4	Ambitious	Broadminded	Forgiving	Broadminded
5	Broadminded	Ambitious	Helpful	Imaginative
6	Courageous	Forgiving	Broadminded	Forgiving

"Honest," "Forgiving," and "Broadminded" appear as most important instrumental values for all groups. "Loving" and "Helpful" are ranked among the most important instrumental values by the entering University of Georgia students, but not by the other two groups. Both groups in the national sample include "Ambitious" and "Courageous" among the top six values while the entering students do not. Unlike the other groups, the 1975-76 group included "Imaginative" among the six most important instrumental values.

Identification of the six least important instrumental values shows more similarities than differences among the groups. These six values are, in rank order, as follows:

Table 8

Median Scores of Instrumental Values on Rokeach Value Survey for Students Entering the Division Compared with Male and Female Subjects in a National Sample with Some College Experience and Females in a National Sample*

		VALUE SURVEY MEDIAN SCORES			
	Instrumental Values	National Sample		Students Entering Division	
		Female	Some College	1974-1975	1975-1976
1.	Ambitious	7.33 (4)	7.79 (5)	9.61 (8)	9.50 (9)
2.	Broadminded	7.64 (5)	7.42 (4)	8.10 (6)	8.13 (4)
3.	Capable	10.10 (12)	9.04 (7)	10.92 (12)	8.93 (8)
4.	Cheerful	9.43 (10)	11.13 (14)	8.29 (7)	8.65 (7)
5.	Clean	8.13 (8)	10.63 (13)	14.10 (16)	10.88 (15)
6.	Courageous	8.06 (6)	6.71 (3)	10.50 (11)	8.50 (6)
7.	Forgiving	6.43 (2)	8.90 (6)	6.25 (4)	8.50 (6)
8.	Helpful	8.07 (7)	9.57 (9)	6.67 (5)	7.50 (1)
9.	Honest	3.21 (1)	3.36 (1)	2.71 (1)	7.94 (3)
10.	Imaginative	16.10 (18)	14.00 (17)	14.30 (17)	8.20 (5)
11.	Independent	10.72 (14)	10.21 (11)	11.88 (14)	9.50 (9)
12.	Intellectual	13.22 (16)	10.29 (12)	13.30 (15)	9.94 (11)
13.	Logical	14.65 (17)	12.10 (16)	11.55 (13)	9.90 (10)
14.	Loving	8.64 (9)	9.68 (10)	2.89 (2)	7.83 (2)

Table 8

(Continued)

		VALUE SURVEY MEDIAN SCORES			
		National Sample		Students Entering Division	
	Instrumental Values	Female	Some College	1974-1975	1975-1976
15.	Obedient	13.08 (15)	14.73 (18)	15.00 (18)	10.60 (13)
16.	Polite	10.71 (13)	11.61 (15)	9.88 (10)	10.70 (14)
17.	Responsible	6.82 (3)	5.85 (2)	6.00 (3)	11.89 (16)
18.	Self-Controlled	9.55 (11)	9.15 (8)	9.61 (8)	10.24 (12)

Numbers in Parentheses Represent Rank-Order of Values of Each Group.

*Rokeach, M. The Nature of Human Values. New York: The Free Press, 1973, pp. 364-391.

Rank	Female	Some College	Entering Students 1974-75	Entering Students 1975-76
18	Imaginative	Obedient	Obedient	Responsible
17	Logical	Imaginative	Imaginative	Clean
16	Intellectual	Logical	Clean	Polite
15	Obedient	Polite	Intellectual	Obedient
14	Independent	Cheerful	Independent	Self-controlled
13	Polite	Clean	Logical	Intellectual

Only the value "Obedient" was considered least important by all groups. Entering students for both years, 1974-75 and 1975-76, considered "Clean," "Obedient," and "Intellectual" to be among least important instrumental values. "Clean" ranked among the six least important values for entering students and for the national sample having some college; the female group ranking and median score, however, are very different from these other two groups. The median score for "Clean" for the female groups was 8.06 (rank of 8) while the scores and rankings for the other groups, Some College Students and Entering Students, were 10.63 (rank of 13); 14.10 (rank of 16); and 10.88 (rank of 15) respectively. The values "Responsible" and "Self-controlled" which the 1975-76 entering students included as least important were different from all other groups. The interpretation of these data is difficult. Rokeach has suggested that values are cognitive representations of needs. Thus, it would follow that variations of value rankings should be related to variations in needs. "A World at Peace" was ranked as one of the highest values for both groups in the national sample in 1968, a time when the United States was involved in the Vietnamese conflict. However, for the students entering the division during 1974-75 and 1975-76, a period when the United States was not involved in any war, "A World at Peace" was not included among the highest values. The investigating committee felt that a cluster analysis of data obtained from the Value Survey might become important in describing values which may be associated with effective teaching, in identifying needs of prospective teachers from a value perspective, and in planning a teacher education program to meet those needs.

The University of Georgia Semantic-Differential Inventory (Form EEL 1975)

The members of the committee agree that the meanings entering students ascribe to certain concepts associated with teaching as a profession should be investigated. It is assumed that a future teacher's behavior may be determined in part by the significance or conceptual meaning placed in particular persons, places, or things in the educational environment. The semantic-differential techniques (Osgood et al., 1975) was selected as the means by which to obtain these data.

The semantic-differential techniques utilizes adjective pairs selected from an empirically tested list which provides extreme points for responses as to how one "feels" when confronted with a particular concept word or phrase. After the subject reads the word or phrase, they respond to it on a seven point scale to indicate which point best represents the meaning that concept holds. Figure 2 presents a sample page from inventory showing the twelve sets of adjectives and the concept "Learning." The remainder of the concepts are listed as column headings in Figure 3 where the findings are summarized.

Figure 2

A Sample Page from the University of Georgia Semantic-Differential Inventory (Form EEL 1975)

Learning

LARGE	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	SMALL
UNPLEASANT	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	PLEASANT
FAST	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	SLOW
DULL	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	SHARP
THIN	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	THICK
HAPPY	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	SAD
WEAK	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	STRONG
GOOD	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	BAD
MOVING	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	STILL
UNFAIR	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	FAIR
PASSIVE	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	ACTIVE
HEAVY	___ 1 :	___ 2 :	___ 3 :	___ 4 :	___ 5 :	___ 6 :	___ 7 :	LIGHT

The concepts were chosen to represent education both as a process (e.g., "Teaching," "Teacher Education at the University of Georgia," and "Pupil Discipline in School") and as a profession (e.g., "School Teachers," "Principal's Role in the Schools," and "Schools for Children in the United States"). Some concepts were selected to describe the attitudes held by students about education as a human enterprise (e.g., "Children" and "Myself"). Ten concepts were expressed in commonly used words to avoid the problems introduced by more technical language (e.g., "Curriculum," "Objectives," and "Teaching Competencies").

Meanings were investigated on the three dimensions recommended by the authors: potency, evaluation, and activity. A high potency rating to a concept usually means that it is perceived as strong and/or powerful; a high evaluation rating usually means it is valued as good, right or beautiful; and a high activity rating usually means that it is perceived as having or requiring a great deal of movement and/or energy.

The semantic-differential in this investigation is a forced choice, three factor, 120 item test (10 concepts times 12 scales) employing the Form II graphic scale method. One sheet of paper is used for each concept, with all judgments elicited successively as shown in Figure 2. Scales representing the same factor are alternated in polarity direction (e.g., unfair - fair but good - bad) to prevent the formation of position preferences. Further, the order of the factors is rotated.

Figures 3 and 4 summarize the mean scores of the entering students, showing that students entering the Division generally ranked the concepts higher in the evaluation and activity dimensions than in the potency dimension. This

Figure 3

Profile of Students Entering the Division in 1974-75
in Terms of Mean Scores Obtained by Semantic Differential
Dimensions of Potency, Evaluation, and Activity
on Ten Selected Concepts in Education

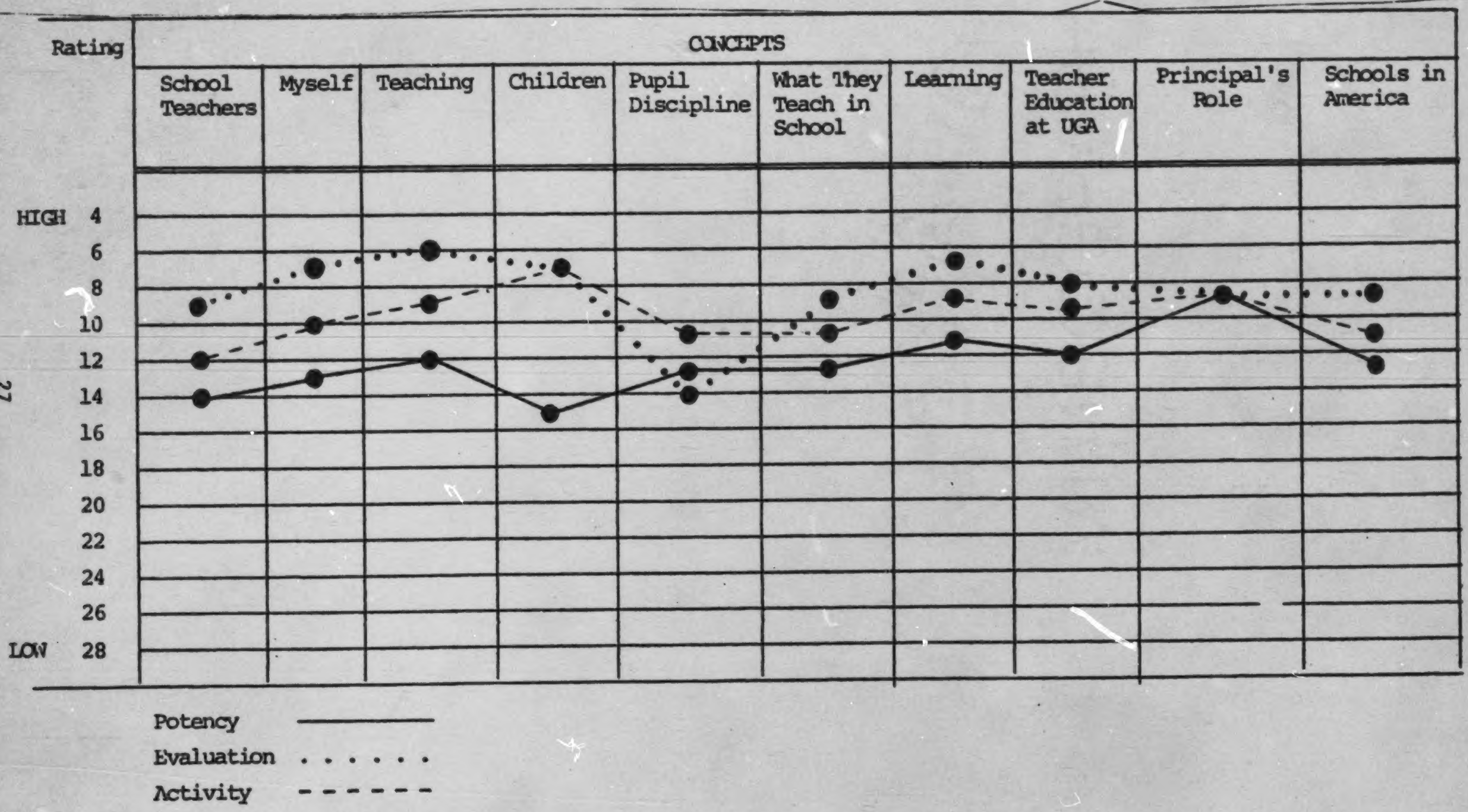
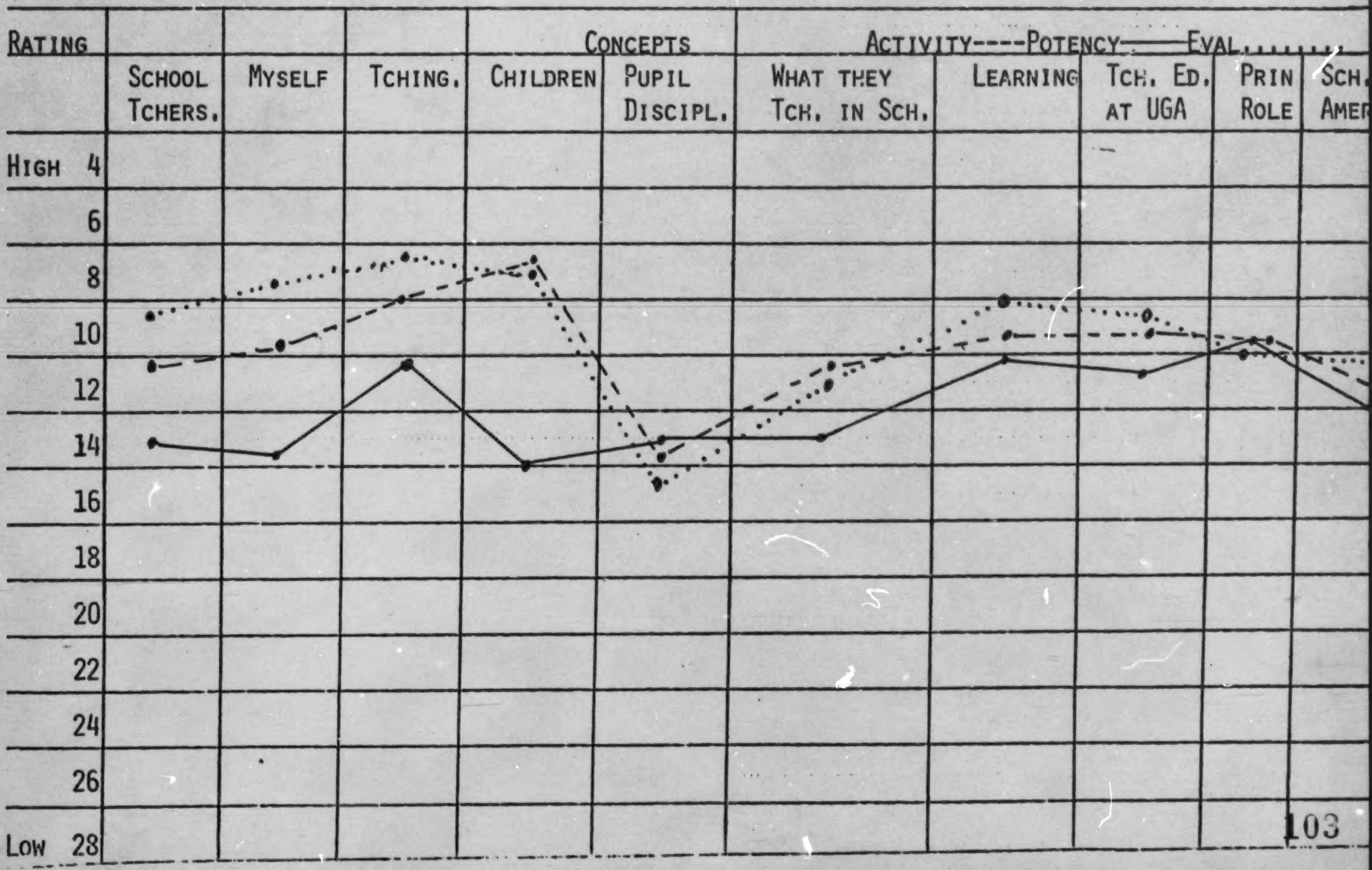


FIGURE 4

PROFILE OF STUDENTS ENTERING THE DIVISION IN 1975-76 IN TERMS OF MEAN SCORES OBTAINED BY SEMANTIC DIFFERENTIAL DIMENSIONS OF POTENCY, EVALUATION, AND ACTIVITY ON TEN SELECTED CONCEPTS IN EDUCATION



may be interpreted to indicate that students perceived educational concepts as relating more to goodness and energy than to influence or authority. This did not hold for "Pupil Discipline in the Schools" or for "Principal's Role in the Schools." Figures 3 and 4 also show that students valued "Myself," "Teaching," "Children," and "Learning" more than the other concepts.

Summary

This paper has presented descriptive information regarding undergraduate students admitted to Elementary Education Division programs in teacher preparation at the early childhood, elementary, and middle school levels. The investigators believe these personal data do not have predictive power at the present time. They may have value for other colleges with comparative data. The real value for this investigation will be in follow-up activities which have been planned in the research design.

Entry data has been obtained from students who entered the program during 1976-77. This data will be presented in a fourth report by the committee.

Additional studies which are in progress include: collecting and reporting data from undergraduate students as they complete the division's programs in teacher preparation comparing entering and exit data; descriptions of each of the programs; and follow-up of graduates after they enter the teaching profession. These studies will be presented in subsequent reports.

References

Brown, B. B. *The Experimental Mind*. New York: Harper and Row, 1968.

Cattell, R. B., Eber, H. W., and Tatsuoka, M. M. *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign: Institute for Personality and Ability Testing, 1970.

Cryan, J. R., Johnson, C. E., and Hawn, H. C. *Second Report of the Committee on Program Research and Evaluation*. Athens, University of Georgia, 1976.

Hawn, H. C., Cryan, J. R., and Johnson, C. E. *Initial Report of the Committee on Program Research and Evaluation*. Athens: University of Georgia, 1975.

Osgood, C., Suci, F., and Tannenbaum, P. *The Measurement of Meaning*. Urbana: University of Illinois Press, 1957.

Rokeach, M. *The Nature of Human Values*. New York: The Free Press, 1973.

Rokeach, M. *The Open and Closed Mind*. New York: Basic Books, 1960.

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ellen Gruber
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

THE EFFECTS OF AN INSERVICE CREATIVITY WORKSHOP ON TEACHERS AND THEIR STUDENTS

Ellen Gruber, Ph.D. and Jeffrey L. Lorentz, Ed.D.
West Georgia College

In recent years a number of educators have moved from an instructional approach which considers only student cognition to an awareness of the need to attend to the affective domain. There is an increasing realization of the need to assist teachers in fostering an affective environment in their classrooms. One approach to this end is the inservice program.

The purpose of the present study was to examine the effects of an Affective and Creative Education inservice program on teacher attitudes and subsequent changes in student attitudes and creativity. Specifically, the study was designed to consider whether a workshop in affect/creativity would result in changes in teacher openness, student creativity and student self-concept.

METHOD

Subjects. The subjects of the present study were eighteen first grade teachers from Carrollton and Carroll County, Georgia who volunteered to participate. Ninety students were obtained by randomly selecting 5 from each of the 18 classrooms. As a condition of participation, each teacher agreed to the administration of student measures both 1 week before the workshop began and 4 weeks after the workshop ended. The teachers also agreed to respond to several measures before and after the workshop.

Instruments. *I Feel - Me Feel* (IFMF) (Yeatts and Bentley, 1970) was given to the 90 students one week before and four weeks after the workshop. This 40-item Likert-type scale (using five faces which ranged from happy to sad rather than numbers) is a self-concept measure appropriate for use with children at this level.

Scoring was based on locally-developed factor keys which were entitled: Academic, Self, Frustration, Femininity, Fun and Independence. Coefficient Alpha reliabilities for these scales ranged from .06 to .92 for the pretest sample.

The *Torrance Tests of Creative Thinking* (TTCT) (Torrance, 1974) were administered along with the IFMF. TTCT is a measure of *creativity* developed over a number of years. A number of studies report realiabilities ranging from .50 to .93 and Torrance (1972) reports nine studies in which TTCT scores were used to predict other student outcomes.

The four subtest scores are fluency, flexibility, originality and elaboration. Figural forms A & B were used.

The *Rokeach Dogmatism Scale* (RDS) (Rokeach, 1960) was administered to each teacher before and after the workshop. Dogmatism as defined by Rokeach is a closed way of thinking which can be associated with any ideology. This scale is designed to measure individual differences in openness or closedness of belief systems. The reliability coefficient (split-half, Spearman-Brown) for the Dogmatism scale, Form E, is .81 based on a group of 80 English colleges. When intercorrelated with Form F among Dogmatism, Authoritarianism, and Ethnocentrism, the R=.62 based on 80 English colleges. Total score is obtained by summing the responses to the 40 items.

A group of specially trained testers administered both the IFMF and TTCT away from the teachers' classrooms.

Treatment. The eighteen 1st grade teacher participated in a two day intensive training program designed to foster affective growth in themselves and subsequently in their students. The workshop focused on the following areas: affective exercises in trust, awareness and communciation, transaction analysis and creativity exercises centered on flexibility. The aim of the workshop was to equip each teacher with skills and a variety of techniques to facilitate the development of an affective curriculum for children. The teachers followed up the workshop via implementation of an affective and creative curriculum for their students.

During the four weeks before the end of the final testing, the teacher's classes, were monitored for one hour once a week to check on the implementation of the techniques taught in the workshop. Trained observers marked a check list in order to obtain additional data on the implementation.

Analysis. The RDS were hand scored and a total pretest, total posttest and difference score recorded for each teacher. A correlated t-test was used to test the difference between pre-RDS and post-RDS. Table 1 presents the results of this test

TABLE 1

RDS Means and Standard Deviations

Test	Mean	Standard Deviation	t
Pre-workshop	135.39	29.01	0.06
Post-workshop	135.05	35.80	

which shows that the difference between pretest and posttest are not significant.

Student IFMF and TTCT subscales were reduced to gains by regressing pretest on posttest and subtracting the obtained posttest score from the predicted ("expected") posttest score and further reducing these gains to classroom means. The 10 scores were then used as a measure of teacher influence on student self-concept and creativity. Teachers were subsequently contrasted on these scores through a series of discriminant analyses (Veldman, 1967). In the first of these, the eighteen teachers were ranked according to their initial (pretest) RDS score and the nine high scoring teachers were contrasted against the nine low scoring teachers. Likewise, the final (posttest) RDS and the difference between pre and post were used in the same manner.

The results of these three analyses are shown in Tables 2 and 3.

TABLE 2

SUMMARY OF 3 DISCRIMINANT ANALYSES

CONTRASTING GROUPS ON IFMF & TTCT

ANALYSIS	WILKS LAMBDA	F	df	χ^2	ρ
1. High vs Low Pre-RDS	0.292	1.700	10,7	--	0.2466
Root 1 (100% of variance)			10	14.787	0.1415
2. High vs Low Post-RDS	0.345	1.317	10,7	--	0.3669
Root 1 (100% of variance)			10	12.698	0.2421
3. High vs Low Change in RDS	0.295	1.674	10,7	--	0.2533
Root 1 (100% of variance)			10	14.653	0.1467

TABLE 3

GROUP MEANS AND UNIVARIATE F-TESTS

FOR 10 STUDENT VARIABLES

Variable	High RDS Group Mean	Low RDS Group Mean	F-ratio df=1,16	P
Analysis 1. Pretest RDS				
IFMF				
1. Fun	48.3678	52.5913	5.1235	0.0360
2. Academic	48.9251	48.9595	0.0005	0.9803
3. Frustration	11.5837	13.4830	3.9879	0.0605
4. Femininity	11.8674	12.4605	1.0411	0.3240
5. Self	22.4306	24.2419	1.7424	0.2032
6. Independence	10.7063	11.0446	0.2299	0.6425
TTCT				
7. Fluency	21.8952	24.1943	0.9449	0.6526
8. Flexibility	12.1718	12.5993	0.0870	0.7686
9. Originality	8.7112	9.1544	0.0261	0.8679
10. Elaboration	16.8456	15.0086	0.6984	0.5796
Analysis 2. Posttest RDS				
IFMF				
1. Fun	49.9014	51.0578	0.2965	0.5993
2. Academic	48.6697	49.2150	0.0586	0.8065
3. Frustration	11.7583	13.3084	2.4520	0.1338
4. Femininity	12.0920	12.2359	0.0578	0.8077
5. Self	23.2718	23.4006	0.0083	0.9258
6. Independence	10.6760	11.0749	0.3213	0.5846
TTCT				
7. Fluency	23.7375	22.3520	0.3311	0.5791
8. Flexibility	13.1045	11.6667	1.0424	0.3237
9. Originality	10.8235	7.0421	2.1533	0.1588
10. Elaboration	17.9931	13.8612	4.2937	0.0523
Analysis 3. Difference RDS				
IFMF				
1. Fun	49.9444	51.0148	0.2538	0.6262
2. Academic	48.9630	48.9216	0.0005	0.9803
3. Frustration	12.2194	12.8473	0.3567	0.5649
4. Femininity	12.1214	12.2065	0.0199	0.8844
5. Self	23.6229	23.0495	0.1595	0.6965
6. Independence	10.6471	11.1038	0.4242	0.5304
TTCT				
7. Fluency	24.0271	22.0624	0.6793	0.5731
8. Flexibility	12.4679	12.3033	0.0129	0.9072
9. Originality	9.8231	8.0425	0.4322	0.5266
10. Elaboration	17.3394	14.5149	1.7554	0.2016

These results indicate that the groups did not differ significantly on any of the measures.

CONCLUSION

The present study was undertaken to study the effects of an affective workshop on teacher openness and on their student's creativity. The results indicate that there has been no significant impact of the workshop on either teachers or their students.

The Affective Education workshop did not lead to significant differences between groups on teacher openness and subsequent creativity measures in children. The workshop may well be effective in changing openness and creativity of children but this study did not demonstrate that.

The length of the affective workshop may be a critical factor. The length of time spent by teachers implementing the affective curriculum in their classroom may be a critical factor. Research should be undertaken in which such factors as size of the workshop, duration of the workshop, and specific activities are varied.

It is possible that long term effects of the affective and creative education inservice program on teacher attitudes and subsequent changes in student attitudes and creativity will be evident over a lengthier time period.

A Research and Evaluation Model for Teacher Education Programs

A Paper Presented at GERA
Atlanta, Georgia, April 28, 1977

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Horace C. Hawn
Charles E. Johnson
John R. Cryan
Judith Reiff
University of Georgia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Horace C. Hawn
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Introduction

Teacher preparation programs have generally been targets of considerable criticism during the past two decades. Criticism has centered on the lack of adequate evaluation procedures for assessing the effectiveness of instructional programs in teacher preparation institutions. Some of the common criticisms and deficiencies found in traditional teacher preparation programs are summarized by Borg, et al., (1970): (a) emphasis is placed on telling rather than doing so that instruction is separate from actual classroom performance; (b) instruction is general rather than sepcific; (c) effective training models are not provided, and when they are, the trainee does not have opportunities to practice the teaching skills he has observed; and (d) effective feedback to the trainee is not provided. One response to the criticism has been a restructuring of traditional approaches to teacher education toward field-based training models.

At the University of Georgia, the College of Education's involvement in and commitment to innovative approaches to teacher education extend from 1968 when it produced a field-centered model. The university currently has a variety of training programs at various stages of implementation which incorporate the field-based approach to teacher education. Providing "real-life" teaching experiences throughout the program sequence, informing students of skills to be acquired, and placing students with effective role models are all central to teacher training at the University of Georgia. However, in the present stage of development, UGA teacher education programs lack a systematic and well-formulated assessment system to enhance both student and program evaluation. Assessment, in this paper, is distinguished from other activities such as measurement and testing, and embraces a variety of methodologies and instrumentation relevant to data-based, educational decision-making.

Program evaluation and student assessment systems have remained largely subjective processes in most programs. Since both training and evaluation activities are an integral part of most data-based innovations in education, establishing a "high-quality" data base and assessment system are priorities in any teacher education program. Without ongoing, coordinated assessment activities to complement training programs, decisions regarding the utility and importance of various approaches are considered meaningless.

The University of Georgia Committee for Program Research and Evaluation (PRECO) has begun an initial investigation of the role that specified program variables and student personal characteristics play in determining program effectiveness and baseline information has been collected on 296 students entering UGA programs since the Fall of 1974. Recognizing the need for a comprehensive assessment system to complement other aspects of the Georgia teacher education model a preliminary evaluation plan has been structured by PRECO.

The belief that an accurate appraisal of entering students is a crucial first step in building a comprehensive assessment system for the program at the University of Georgia has guided PRECO efforts to date. It is expected that continuous assessments of students as they progress through the training sequence, as well as "on-the-job" assessments of program graduates will become integral parts of a more comprehensive program assessment and evaluation model.

Intended Outcomes

To deal effectively with the problems that have been identified in teacher training, the major thrust of this effort will be the design and implementation of a comprehensive assessment system to promote more effective instruction through the use of feedback to students, instructors, and program developers. To date the efforts of PRECO have been directed toward accomplishing the following: (a) selecting and refining instruments appropriate for determining the personal characteristics of students currently entering the teacher education sequences; (b) establishing formal and informal procedures and schedules for collecting information from students, and for the processing, storing, and analysis of such data; (c) reviewing instrumentation appropriate for the evaluation of preservice teacher behavior; (d) reviewing instrumentation appropriate for the evaluation of elementary teacher preservice education programs; (e) analyzing criterion referenced instruments developed to date and their usefulness in evaluating aspects of the UGA teacher education programs.

With progress being made toward the implementation of a comprehensive assessment system and program evaluation model, the immediate objectives of PRECO and a subsequent comprehensive proposal are: (a) utilizing personal characteristics information on entering students to predict success in the teacher education program; (b) utilizing personal characteristics data on entering students for teacher education program counseling; (c) developing a set of specific objectives and criterion measures to be used as one form of program criteria within the comprehensive assessment schema; and (d) refining and further developing the program evaluation model to reflect a variety of assessment inputs related to long-range project goals, including the evaluation of elements in the instructional sequence.

teachers. In addition, other criterion variables such as grade point average, NTE scores, students teaching performance records, etc., will be added to the data base as student "outcomes" essential for both monitoring students, and assessing the effectiveness of components of the total training program. A large amount of effort will be spent in developing psychometrically sound instrumentation for measuring multiple aspects of student performance.

The third procedural category, program evaluation, will delineate parameters of a short range and longitudinal evaluation model for both the instructional program and its "participants", and the necessary research investigations that will be undertaken to implement the evaluation design. Essential to the design are analyses leading to a better understanding of the instructional program and its various components on students' acquisition of knowledges, skills, and attitudes deemed essential for beginning teachers. Instruments currently being administered to all students entering the program will be revised contingent upon analyses of their measurement properties and their ability to predict meaningful program criteria once these are established.

In addition to these analyses, participants (students) in the program will be evaluated within the comprehensive assessment design. Specific studies will be undertaken to better understand personal characteristics of entering and exiting students, the types of training each has received, and how specific dimensions of the instructional program interact with these variables. These investigations should facilitate the future selection, counseling, and monitoring of students as they progress through program sequences. It is expected that the longitudinal program evaluation will include a series of follow-up investigations of graduates after an initial teaching year in the public schools.

A final procedural category for accomplishing project objectives in concerned with program revision and improvement. Evaluative data concerning

The long-range objectives have been formulated to improve the quality of instruction, teacher education programs, and prospective beginning teachers exiting the institution. It is expected that the efforts described here will eventuate in a transportable program training and evaluation model that is data-based, has been validated in practice, and exhibits wide flexibility in meeting student personal characteristics and training needs. It is expected that a data base having relevance for improving teacher preparation through adequate assessment and evaluation models will be established within the College of Education at UGA.

Procedures

The general procedures for accomplishing both the immediate and long-range objectives fall within four general categories: (1) the data base, (2) criterion measures, (3) program evaluation, and (4) program revision and improvement.

For the first category, the data base, a more efficient system of data collection, keypunching, and information storage and retrieval will be established for currently administered instruments. Supplemental background information on entering and exiting students will be added to the data base where this will enhance accomplishment of the terminal objectives.

The second category of procedures centers on criterion measures. Efforts will be undertaken to develop and refine criterion measures to be administered to students going through the educational sequence at several different points in the program. These measures will consist of a variety of "low" and "high" inference techniques ranging from observation instruments for monitoring the acquisition of specific teaching skills to rating scales derived from a professionally verified list of behaviors considered essential for beginning aspects of the educational program will be obtained from all personnel involved in training...students, instructors, field supervisors, and program administrators. These data will be derived from a variety of methodologies emphasizing the use of formative and summative evaluation instruments and techniques. It is expected that data collected will facilitate the structuring of content and training procedures. When integrated with evaluative data from other sources explicated above, a data feedback system will be established leading to continuous program revision and improvement.

Instrumentation

The following is a list of information currently being collected from students entering teacher education programs at the University of Georgia:

A. Entering Student Information Inventory (biographic and demographic data)

B. Sixteen Personality Factor Questionnaire (16PF) (Raymond B. Cattell, 1967)

C. California Test of Mental Maturity (Form S, Level 5) (McGraw Hill, Inc., 1963)

D. Personal Beliefs Inventory (Bob Burton Brown, 1968)

E. Teacher Practices Inventory (Bob Burton Brown, 1968)

F. Dogmatism Scale (D Scale) (Milton Rokeach, 1960)

G. Value Survey (Milton Rokeach, 1967)

F. University of Georgia Semantic-Differential Inventory (10 teacher education concepts)

Reference

Borg, W. R., and M. L. Kelley, P. Langer. The Mini-course, a Micro-teaching Approach to Teacher Education, Beverly Hills: Macmillan Educational Services, 1970.

EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AFFECTIVE CORRELATES OF CLASSROOM BEHAVIOR AMONG PRIMARY-LEVEL STUDENTS*

Jeffrey L. Lorentz
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Jeffrey L. Lorentz
West Georgia College

The present paper describes steps taken in the development of scoring keys for a student affect measure and presents some of the relationships between this measure and measures of teacher behavior.

The data were collected as part of the Carroll County, Georgia, Competency Based Teacher Certification Project. The rationale for and the design of this project were reported in an AERA paper (Lorentz, 1976) which includes a complete description of all the instruments used and a thorough discussion of the procedures.

The instrument "I Feel-Me Feel" (IFMF) (Yeatts and Bentley, 1970) was selected for use with students in grades 1-3. This 40 item Likert-type scale is a self-concept measure appropriate for use with children at this level, and had been used in other studies of this type.

In September, 1974, three alternative strategies for scoring the instrument were available to CBTC staff. These included a set of composites used by the Soars (Soar and Soar, 1973), an alternate set of composites of unknown origin, and a CBTC developed set of composites.

The keys used by the Soars contained a number of items in more than one subscale. This, it was felt, would result in highly correlated subscales. The other keys available contained fewer than half of the items and only one scale contained more than three items. Therefore, CBTC staff chose to develop its own subscales and proceeded to do so in the following manner:

The instrument was administered during the Fall, 1974, to all students in grades 1-3, in project classrooms. After eliminating the responses of those children who were obviously unable to respond (papers with marks in regular 1-2-3-4-5 patterns, scribbles, etc.) 483 records were available for further analysis. IFMF was subsequently readministered during the Spring, 1975, in project classrooms to 465 students. Each student's record form was keypunched and verified.

All 40 items were factor analyzed (Guertin and Bailey, 1970), using the entire Fall population of 483 students. The principal axes method, with multiple R^2 as the communality estimate, was used to obtain the initial solution. A series of orthogonal rotations were then performed in which two through six factors were rotated. The six factor solution accounted for 37.4% of the total score variance, or 83.4% of all common variance; this six factor solution was selected for interpretation. The rotated factor matrix is shown in Table I.

---

* Presented at the Annual Meeting, Georgia Educational Research Association, Atlanta. April 28, 1977.

Insert Table I about here

Items which loaded highest on each factor were examined for a common psychological dimension and each factor was given a title which reflected this dimension. Items loading on Factor 1 seemed to relate to a child's perception of school-related activities and this factor was called "Academic". Factor 2 appeared to be a measure of the child's self-concept and was called "Self". Factor 3 contained two items (related to the principal and to Arithmetic) which may or may not be in accord with its "Frustration" label, although the other two items clearly support this label. The three items loading on Factor 4 are (at least for young children) a measure of "Femininity". While some of the Factor 5 items were clearly school-related, the underlying thread seemed to be a child's perception of "Fun". Finally, Factor 6 seems clearly to be a measure of "Independence".

The interpretation of these factors is, of course, open to discussion; however, Crouthamel (1975) identified five factors which he called Fun, Academic, Femininity, Self Assertion, and Detatchment. There is enough similarity between the six CBTC-derived factors and Crouthamel's five to confirm the appropriateness of the factor labels.

Coefficient alpha reliabilities computed for the six factor keys were high enough to support the decision to use these keys as indices of student growth in the CBTC project. Total raw scores were obtained on each factor key for each student, using unit weights in the scoring procedure. Means and standard deviations and reliabilities for both the Fall and Spring groups are shown in Table II.

Insert Table II about here

One measure of the contribution of an item to the total score is the item-total correlation; since the item is contained in the total score, a better measure of its contribution is the item-total correlation, corrected for the effect of that item (Nunnally, 1967). These correlations, for the Fall data, are presented in Table III.

Insert Table III about here

With the exception of the Frustration Factor, and one item in the Fun Factor, all corrected item-total correlations are greater than .40. The remaining items correlate with the total at an adequate level (a corrected correlation greater than approximately .10 is significant at the .001 level). Again, further confirmation of the use of these keys as measures of student growth.

There is a generally held suspicion that students' attitudes toward school may change adversely over the school year: "Academic" down, "Frustration" up. However, while the mean "Academic" score decreases one point, the remaining means remain about the same. Standard deviations and reliabilities become smaller for the Spring sample which suggests that the pattern of responses is different in the Spring.

Students in each of the 18 classes tested with IFMF were matched and adjusted posttest means (i.e., regressed gains) were computed for each Factor using the initial test score and socioeconomic status (SES) as covariates. These regressed gains are the average of the difference between the predicted Spring score and the actual Spring score of each student in a class. They are measures of student attitudes which are independent of Fall attitude and SES. A paper discussing the rationale and computational procedures was presented at AERA (Lorentz, 1977).

Estimates of the intraclass reliability of each Factor score was obtained using the formula

$$r = \frac{MS_b - MS_w}{MS_b}$$

where: MS_b = mean square between classes
MS_w = mean square within classes
from the Analysis of Covariance

The adjusted means and intraclass reliability coefficients for the 18 classes are presented in Table IV.

Insert Table IV about here

Sixty-two measures of teacher competency were derived from five objective observation instruments applied over a year-long period in the CBTC classrooms (Medley, 1977). These "competency scores" for seventeen of the classrooms (grades 1-3) were correlated with the mean IFMF residuals. The significant correlations ($p < .05$) are presented in Table V.

Insert Table V about here

These reveal some interesting results, some contrary to expectation. It appears, for example, that setting appropriate goals and objectives may engender adverse student attitudes. Students able to follow directions generally have higher "Frustration" and more "Fun" than those who don't. And students who ask questions have better self-concept and more fun. Independence, not surprisingly, is related to individualized treatment by the teacher. Finally, it is interesting to speculate why confidence in the teacher is apparently low among girls.

The correlations of IFMF residuals with achievement residuals revealed that IFMF is generally uncorrelated with achievement. The only significant correlation was .91 between IFMF Frustration and achievement in Mathematics Concepts. This is reasonable given the item on that key related to Arithmetic.

In summary, the use of an affective measure among primary-level students can produce stable indices of student attitude. These indices can serve as useful outcome measures in process-product studies.

REFERENCES

Crouthamel, W.W. *"I Feel-Me Feel" Self Concept Appraisal Technical Manual*. Metropolitan Cooperative Educational Service Agency, 1975 (Mimeographed).

Guertin, W.H. & Bailey, J.P., Jr. *Introduction to Modern Factor Analysis*. Ann Arbor: Edwards Brothers, Inc., 1970.

Lorentz, Jeffrey L. *The Design of a Study to Validate Teacher Competencies in Terms of Pupil Growth*. A paper presented at the 1976 Annual Meeting, American Educational Research Association, San Francisco, California, April, 1976. (ERIC ED No. 126-152)

Lorentz, Jeffrey L. *The Development of Measures of Teacher Effectiveness from Multiple Measures of Student Growth*. A paper presented at the 1977 Annual Meeting, American Educational Research Association, New York, New York, April, 1977.

Medley, D.M. *An Approach to the Definition and Measurement of Teacher Competency*. A paper presented at the 1977 Annual Meeting, American Educational Research Association, New York, New York, April, 1977.

Nunnally, J.C. *Psychometric Theory.* New York: McGraw-Hill, Inc., 1967.

Soar, R.S. & Soar, R.M. Classroom Behavior, Pupil Characteristics, and Pupil Growth for the School Year and for the Summer. Gainesville, Florida: Institute for Development of Human Resources, University of Florida, 1973. JSAS *Catalog of Selected Documents in Psychology*, 1975, *5*, 200. (Ms. No. 873)

Yeatts, P.P. & Bentley, E.L., Jr. *I Feel-Me Feel Self Concept Appraisal*. (Privately Published), 1970.

ITEM	FACTOR 1 ACADEMIC	 2 SELF	 3 FRUSTRATION	 4 FEMININITY	 5 FUN	 6 INDEPENDENCE
1. Walking with the teacher	51*	11	-03	06	17	-05
2. Building things with blocks	50*	23	08	-09	15	11
3. Doing things for the first time	50*	10	15	17	07	23
4. School	48*	01	10	11	18	28
5. When the teacher talks	41*	26	14	10	15	18
6. Counting to ten or more	52*	10	16	09	11	04
7. Singing	42*	22	04	26	16	15
8. Reading a new book	56*	20	12	03	22	00
9. Doing things by myself	33	18	23	08	03	36*
10. Going to school	46*	03	05	16	18	31
11. When I talk to the teacher	39*	30	12	14	23	06
12. When I get mad	04	10	37*	01	-01	01
13. My clothes	42*	27	08	10	36	-05
14. Dogs and cats	28	28	01	16	29*	26
15. When I draw	39*	27	02	16	35	04
16. Flowers	34	21	09	41*	32	11
17. Looking in a mirror	31	38*	30	13	20	13
18. Running	29	51*	18	-01	22	12
19. When I paint	23	30	01	11	43*	20
20. Writing	28	39*	23	19	16	17
21. Children	32	27	-02	22	37*	10
22. My hair	23	55*	06	22	20	11
23. Big people	15	31*	28	18	26	12
24. Building things by myself	07	29	16	02	26	41*
25. When I play	29	38	-11	18	47*	12
26. Reading	22	30	17	23	28	33*
27. When I dance	14	25	20	51*	15	09
28. Throwing a ball	21	42*	21	-02	37	13
29. Eating lunch at school	23	33	17	22	39*	10
30. My class at school	17	18	11	14	43*	33
31. Going down a slide	26	12	17	21	61*	03
32. Doing things I've done before	09	15	42*	09	33	21
33. Listening to stories	16	13	18	23	53*	13
34. Playing games	26	20	18	17	52*	05
35. Arithmetic	12	01	35*	28	23	16
36. Boys	18	17	25	-15	30*	17
37. Playing with girls	09	03	07	57*	27	04
38. The teacher	24	11	21	20	47*	26
39. Books	21	21	30	21	43*	10
40. The principal's office	06	-01	45*	09	11	04

* items included in key

TABLE II

MEANS, STANDARD DEVIATION, AND COEFFICIENT ALPHA RELIABILITIES SIX FACTOR KEYS AND TOTAL SCORE

Factor Key	# of Items	Fall, 1974 Mean/Std.Dev/Alpha n = 483			Spring, 1975 Mean/Std.Dev/Alpha n = 465		
1. Academic	12	48.7	9.93	.84	47.6	7.88	.71
2. Self	6	23.2	5.97	.78	23.1	4.92	.64
3. Frustration	4	11.3	4.08	.54	11.4	3.82	.45
4. Femininity	3	11.6	3.52	.64	11.8	3.32	.56
5. Fun	12	50.2	10.13	.86	50.8	7.82	.74
6. Independence	3	10.6	3.58	.60	10.8	3.01	.31
TOTAL	40	155.4	30.22	.93	155.4	22.72	.88

TABLE III

ITEM-SUBSCORE CORRELATIONS
Fall Sample (n = 483)

Item Number	Item-Subscore Correlations	Corrected Correlations
	Factor 1 - Academic	
1	56	47
2	61	50
3	61	51
4	60	50
5	60	49
6	60	50
7	60	50
8	66	57
10	59	49
11	60	49
13	62	53
15	60	51
	Factor 2 - Self	
17	71	56
18	72	56
20	67	51
22	70	53
23	64	44
28	69	54
	Factor 3 - Frustration	
12	52	24
32	66	35
35	68	34
40	71	37

Item Number	Item-Subscore Correlations	Corrected Correlations
	Factor 4 - Femininity	
16	69	41
27	80	50
37	80	45
	Factor 5 - Fun	
14	60	50
19	62	54
21	62	53
25	68	60
29	65	57
30	62	52
31	71	65
33	66	57
34	67	60
36	50	36
38	66	58
39	67	59
	Factor 6 - Independence	
9	75	41
24	77	43
26	70	38

TABLE IV

MEAN REGRESSED GAINS AND INTRA-CLASS RELIABILITIES

Factor		Mean Regressed Gains	Intra-class Reliability
1.	Academic	48.6	.38
2.	Self	23.4	.59
3.	Frustration	11.6	.52
4.	Femininity	12.0	.58
5.	Fun	51.5	.59
6.	Independence	11.1	.37

TABLE V

SIGNIFICANT ($p < .05$) CORRELATIONS BETWEEN MEASURES OF COMPETENCY AND I FEEL-ME FEEL RESIDUALS

COMPETENCY		IFMF FACTORS					
		1 Aca	2 Self	3 Frus	4 Fem	5 Fun	6 Ind
2A-T	Selects goals and objectives appropriate to pupil need	-49	-61			-79	-67
2C-T	Gathers multi-level materials			52			
2B/C-S	Actively involved, working on-task			49			
3A-T	Gives clear explicit directions					-56	
4C-S	Able to follow directions on-task			52		50	
5B-T	Demonstrates flexibility in classroom management practices						52
5A-S	Students ask questions		51			51	
6B-T	Recognizes and treats individual student behavior						54
6D-T	Accepts necessity of dealing with individual students						52
8E-S	Evidence of importance as class member, group involvement			66			49
9D-S	Evidence of confidence in teacher				-51		

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Owen Scott

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The Stability of Students' Appraisals of College Instruction: A Further Look

Owen Scott, Alex F. Perrodin and Carl Schnittjer
University of Georgia, Athens, Georgia 30602

Running Head: Students' Appraisals of College Instruction

Abstract

The authors question the use of the appraisal-reappraisal correlation coefficient as an appropriate measure of the stability of students' appraisals of college instruction, if these are to be used for administrative purposes. In lieu, two other measures, estimated Kappa ($\hat{k}$) and Lambda (λ), are proposed. Illustrative data are presented highlighting differences in the interpretation of these measures and some additional suggestions are made relevant to the administrative use of students' appraisals of instruction.

The Stability of Students' Appraisals of College Instruction: A Further Look

From time to time since 1954, if not earlier, research has been published reporting the stability over time of students' appraisals of college instruction (see e.g., the excellent review of such studies by Costin, et al. (1971). In most such studies the basic data were individual responses requiring overall or "global" appraisal of the classroom instruction. Moreover, the measure of stability reported was the appraisal-re-appraisal correlation coefficient. These coefficients generally ranged from the .50s to the high .80s with these results cited as evidence of stability. The time interval ranged from a few weeks in most of the studies to a year or more in a few.

More recently, a different approach to the measurement of stability has been reported (Frey, 1976). Frey divided each of seven college classes into random halves with one half of each class appraising instruction during the first week of next quarter. Appraisals, solicited by mail, were obtained from 68% of the end-of-quarter group and from 70% of the delayed appraisals group. The basic data were individual responses to which Frey applied seven 2 (time of data collection) x 7 (number of classes) analyses of variance (ANOVAs), one for each of seven appraisal items, the responses to which were scored on a 1-5 scale. In each ANOVA Frey reported the time effect as not statistically significant (the significance level was not reported but the computed F statistic in each ANOVA was very close to E[F]). These results constituted Frey's evidence of the stability of the appraisals.

Of course, the F-statistics used were, in each ANOVA, testing the hypothesis that the obtained difference between the appraisal mean and the reappraisal mean could reasonably have been obtained from a hypothetical population of instructors and classes in which the means are the same.

Although Frey's study and those reviewed by Costin, et al., are relevant in a general sense to the question of the stability over time of students' appraisals of instruction, they are not very helpful to the administrator who proposes to use such appraisals as partial evidence to support decisions concerning tenure, merit salary raises, and promotion. Decisions pertaining to tenure, merit raises and promotions are usually made in an economy of scarcity. It is not often the case that the administrator can approve tenure, merit raises and promotions for all of those considered, but only for the "most deserving". The phrase "most deserving" implies classifications and comparisons based on these classifications. A direct way of using students' appraisals to make such comparisons is to use a frequency distribution of class mean appraisals for the instructors to be compared and to use measures of relative positions in the distribution as the bases for the comparisons. As an example, an instructor could be classified as "most deserving" in terms of students' appraisals if the class mean for this instructor is in the highest fourth of the distribution. Similarly, a class mean in the lowest fourth could classify an instructor as "least deserving" in terms of students' appraisals. In this context, what is needed as a measure of stability is one which estimates the stability of the classifications from appraisal to reappraisal. Moreover, it should measure the stability of the classifications for these particular instructors rather than estimate the stability of classifications for some hypothetical population of instructors and classes. Neither the correlational data used in most previous studies nor the ANOVA data reported by Frey satisfy both of these requirements.

Two measures which do satisfy them are estimated Kappa ($\hat{k}$) and Lambda (λ). If appraisal and reappraisal means have been obtained and classified (e.g., as in the highest fourth, middle half, or lowest fourth of the respective distributions), $\hat{k}$ is the proportion of the agreement in the two sets of classifications over and above that which can be expected by chance (Cohen, 1972; Swaminatha et al., 1974). $\hat{k}$ can be negative (of course, a negative $\hat{k}$ is prima facie evidence of instability) and has a maximum value of +1 if the two sets of classifications are in perfect agreement. λ is a measure of the accuracy of predicted classifications if knowledge of initial classifications is used as the basis for predicting subsequent classification (Hays, 1973, pp. 745-749). It is therefore, an index of predictive association between the two sets of classifications and is interpreted somewhat like an interpretation of a correlation coefficient. It is the proportional reduction in the probability of predictive errors in reappraisal classification when appraisal classifications are used as the bases for the predictions. λ can range in value from 0 to 1; the larger its value, the more stable the classifications.

Purposes

The purposes of the present study were to obtain student appraisal and reappraisal data for a group of college instructors, to compute and compare r, $\hat{k}$, and λ for selected appraisal items and to make some suggestions which may aid college administrators who are using students' appraisals of instruction in decisions relating to tenure, merit salary raises and promotions.

Method

Instrument. The appraisal form used was the University of Georgia Faculty-Course Evaluation Form containing eight items asking for personal information, 26 items requiring appraisal of specific aspects of instruction and two "global" items, one asking for appraisal of the value of the course and the other, for appraisal of the instructor's teaching ability. Responses to each item are structured and scored on a 1-5 scale.

Data Collection. End-of-quarter appraisals were obtained from the 540 students enrolled in the classes of 34 instructors who volunteered to have their students make appraisals. Approximately three weeks later the students were contacted by mail and again asked to appraise the instruction. To preserve the anonymity of responses and to make possible, also, the pairing of each immediate and delayed appraisal, the authors instructed students to write on their appraisal forms their mothers' maiden names. Since these names were nowhere on record on campus, students were assured of anonymity. Also, on separate cards, unassociated in any way with their appraisal forms, students were asked to write their names and mailing addresses. The authors kept these cards separate by classes and used them to contact the students by mail asking for a second appraisal. Since the origin of the mailed material was not identified, anonymity was again preserved. Of the 540 students contacted, 281 mailed in second appraisals.

In 23 of the 34 classes more than five students responded a second time. The 245 paired responses for these 23 classes comprised the data. (see Table 1).

Insert Table 1 about here

Data Analysis. Two appraisal-reappraisal correlation coefficients (r's) were computed for each of the 28 items; one, using individual responses and one, mean responses for each of the 23 classes. For the two "global" items and for three other items, $\hat{k}$ and λ were also computed. The latter three items were selected as typical of items with r's ranging from moderately low to moderately high.

Results

Table 2 summarizes the correlational data. Twenty-three of the 28 r's computed using individual responses are .50 or above while 26 of the 28 r's based on means are .60 or above. Thus, according to the correlation criterion used in most previous studies, most items are satisfactory in terms of response stability. In fact, however, these data are misleading if one is concerned with the stability of the classifications of means. Table 3 lists for the five items selected, r, $\hat{k}$ and λ.

Insert Table 2 about here

Insert Table 3 about here

Examination of these statistics reveals a number of interesting facts. Item 31, quite acceptable according to the r criterion, is clearly unacceptable according to both the $\hat{k}$ and the λ criterion. The close to zero and zero values of the latter two statistics classify Item 31 as entirely unsatisfactory. Differences such as these among the three statistics are not unique to Item 31 but can occur with any item. A second fact gleaned from Table 3 is the fact that the two items, Items 35 and 36, are almost identical with respect to correlation and yet differ markedly with respect to $\hat{k}$ and λ. Again, this state of affairs is not unique to these two items but can occur with any two items. A third fact illustrated in Table 3 is that $\hat{k}$ and λ measure different aspects of classification stability. For example, if these two statistics are compared for Items 33 and 36, $\hat{k}$ indicates that classification of the means is reasonably stable for both items, with Item 33 having a slight edge. With respect to λ, however, the items are dissimilar with Item 36 superior to Item 33.

Discussion

The thesis of this study is that although concern for the stability of students' appraisals of college instruction is a legitimate one, previous studies have focused on measures of stability which are misleading to the administrator who plans to use the appraisals as one basis for making decisions pertaining to tenure, merit salary raises and promotion. In the authors' judgment the basic data should be class means rather than individual appraisals and some measure of the stability of the classification of the means is the appropriate measure of stability. Two appropriate measures of stability are $\hat{k}$ and λ, and both should be used to measure stability.

Assuming that the caveats of such specialists as Doyle (1975), Frey (1976) and Miller (1972) are heeded with respect to inventory content and data collection procedures, a critical question remaining for the administrator is how to group instructors for purposes of comparison. There is research to suggest that course level (e.g., lower division undergraduate versus upper division undergraduate), course content (e.g., art, history, mathematics), class size, class location (on or off campus), reason for selection (elected vs. required), academic rank of instructor and over-all reputation of the instructor are variables which may influence students' appraisals to a greater extent than students' characteristics (Kulik & McKeachie, 1975, p. 215). With respect to students' characteristics, in one comprehensive study using a variety of characteristics, these accounted for about 21% of the variance in the appraisals (Kulik & McKeachie, 1975, p. 215). Of these student characteristics, perhaps the most important is "the general disposition of the student toward instructors and courses in general" (Kulik & McKeachie, 1975, p. 215). Obviously, administrators cannot group instructors in terms of the characteristics of their students; however, the findings cited above do suggest that instructors should be grouped by course content, course level, class size, class location (on or off campus) and course type (elected or required).

Another troublesome question for the administrator arises in the usual situation in which an instructor has taught more than one class. There is evidence to indicate that class mean appraisals may vary for the same instructor teaching different sections of the same course and may vary even more so for the same instructor teaching different courses. One procedure which may be adopted is to classify each mean in its appropriate distribution. In the event of conflicting classifications, the administrator will have to exercise judgment, taking into account the special circumstances pertaining to each class, but may find it helpful to work out in advance some "rules of thumb" to use in considering such conflicts.

Although some specialists in students' appraisals of instruction have reservations concerning their use for administrative purposes (e.g., Dressel, 1976, Ch. 15), the reality is that they are being so used and will probably continue to be. Under these circumstances, it is the authors' hope that the information presented here may be of some help.

TABLE 1

Number of Paired Appraisals in Each of the 23 Classes Used in This Study

Class No.	n	Class No.	n
1	7	13	7
2	11	14	14
3	7	15	18
4	7	16	11
5	11	17	6
6	9	18	12
7	8	19	9
8	9	20	8
9	11	21	14
10	8	22	13
11	23	23	13
12	9		

TABLE 2

Appraisal-Reappraisal Correlation Coefficients for the Individual Items on the University of Georgia Faculty - Course Evaluation Form

Sizes of r	Number of r's[a]	Number of r's[b]
.900 - .949	---	1
.850 - .899	---	8[c]
.800 - .849	---	5
.750 - .799	---	5
.700 - .749	2	3
.650 - .699	3	3
.600 - .649	9	1
.550 - .599	4	---
.500 - .549	5	---
.450 - .499	3	---
.400 - .449	2	---
.350 - .399	---	2

a. These r's were computed using individual responses. n = 281.

b. These r's were computed using the 23 paired class means.

c. The r's for the two items requiring global approval of the value of the course (#35) and of the teaching ability of the instructor (#36) were in this interval; .837 for #35 and .828 for #36.

TABLE 3

Comparisons Among Three Reliability Measures for Five Items Selected from the University of Georgia Faculty - Course Evaluation Form

Item	r^1	$\hat{k}$	λ
9. The instructor was willing to give individual assistance outside of class.	.39	-.16	.08
31. The instructor explained text materials that were confusing to students.	.64	.04	0.00
33. The instructor seemed well informed about the material presented.	.78	.76	.36
35. How would you rate the over-all value of this course?	.84	.43	.27
36. How would you rate the teaching ability of this instructor?	.83	.65	.54

1. Based on means (n - 23).

References

Cohen, J. Weighted Chi Squares: An extension of the Koppen Method. Educational and Psychological Measurement, 1972, 32, 61-74.

Costin, F., Greenough, W. T. & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 5, 511-35.

Doyle, K. O. Jr. Student Evaluations of Instruction. Lexington, Mass.: D. C. Heath & Co., 1975.

Dressel, P. L. Handbook of Academic Evaluation. San Francisco: Jossey-Bass, Inc., 1976.

Frey, P. W. Validity of instructional ratings. Does timing matter? Journal of Higher Education, 1976, 3, 327-36.

Hays, W. L. Statistics for the Social Sciences, 2nd Ed. New York: Holt, Rinehart & Winston, 1973.

Kulik, J. A. & McKeachie, W. J. The evaluation of teachers in higher education. In Fred N. Kerlinger (Ed.) Review of Research in Education, (Vol. 3), Itasca, Ill., 1975.

Miller, R. I. Evaluating Faculty Performance. San Francisco: Jossey-Bass, Inc., 1972.

Swaminatha, H., Hambleton, R. K., & Algena, I. I. Reliability of criterion-referenced tests: A decision-theoretic formulation. Journal of Educational Measurement, 1974, 11, 263-267.

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Margaret B. Weber

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

An Examination of the Bilevel Dimensionality of Lower and Higher Mental Processes in Probability Achievement*

Margaret B. Weber
Division of Educational Studies
Emory University

Presented at the Georgia Educational Research Association
Annual Meeting, April 28, 1977

*A longer, more detailed description of this instrument is included in Levin (1975). Copies are found in Weber (1976).

Abstract

The purpose of this study was to examine the bilevel dimensionality of probability achievement as operationalized by Levin (1975) using three psychometric procedures: Rasch latent trait analysis, classical item analysis and principal component factor analysis. Evaluating the fit mean square values from the Rasch analysis, it was found that seven items exceed the criterion fit for the unilevel latent trait model. The factor structure lends support to the concept of unidimensionality. Further, the point biserials indicate increased homogeneity among only Rasch fitting items when compared to the point biserials of the original test.

An Examination of the Bilevel Dimensionality of Lower and Higher Mental Processes in Probability Achievement

Too often the validity of non-standardized instruments used in educational research is based only on estimates of content validity. However, as grounds for judging the extent to which test items measure the desired trait are not always clear and may be incorrect, the conclusions reached in these research efforts may be faulty. In a recent study of the effects of process-oriented practice, probability achievement was characterized as multidimensional (Levin, 1975). Two major dimensions were identified: lower mental processes (LMP), i.e., rules knowledge, and higher mental processes (HMP), i.e., application (e.g., Bloom, et al, 1956). Evidence for this definition derived from estimates of content validity and from examination of response patterns. However, no test of the bilevel dimensional achievement measure was reported. The purpose of this study was to examine the bilevel dimensionality of probability achievement as operationalized by Levin (1975) through a variety of psychometric procedures.

Operating under Levin's assumption that the test under examination represented both an LMP and an HMP dimension, the item responses were analyzed to determine if they represented more than one dimension. The items were first fitted to the Rasch latent trait model. Classical item analysis procedures were used to estimate the reliability of the test. In addition, a principal component factor analysis procedure was used to test the multidimensionality of this test. Finally, both classical and factor analytic procedures were used to examine the test as redefined based on the Rasch analysis.

Instrumentation

The original instrument was a sixty-three item probability achievement test of traditional form. The items consisted of statements of conditions followed by a question. The solution required either a short written response or a calculation. The first twenty-nine items had been designed to measure knowledge of rules (e.g., Bloom, et al, 1956). These items define the lower mental process (LMP) dimension of probability achievement. The remaining thirty-four items had been designed to measure application of rules (e.g., Bloom, et al, 1956). These items define the higher mental process (HMP) dimension of probability achievement.

The analyses were performed on a total of forty-seven items. Sixteen items were omitted because they were too difficult for the subjects in the calibration sample. These omitted items were distributed throughout the last thirty-four items suggesting that something other than speed or location contamination accounted for the high proportion of incorrect answers among the HMP items.

Results and Discussion

Evaluating the fit mean square values from the Rasch analyses as F-ratios (critical value = 2.21), one finds that items 2, 12, 19, 25, 27, 30 and 32 exceed the criterion fit for the unidimensional latent trait model (see Table 1). The remaining forty items then

Insert Table 1 About Here

define a unidimensional trait that may be labeled "probability achievement."

Gulliksen (1950) reports that if empirical analysis of a test yields one common factor, the point biserials should be relatively homogeneous. One can expand upon this to suggest that items which fit the Rasch latent trait model should yield relatively homogeneous point biserial correlation coefficients. As indicated in Table 1, the point biserial values for items fitting the Rasch model range from .03 to .59. The KR_{20} for the original 48 LMP-HMP test is .80 (see Table 2). If the test is redefined according to the fit mean square criterion, the homogeneity of the revised test should increase. Thus, the reliability of the test should also increase. And, as reported in Table 2, the KR_{20} for the revised LMP-HMP test does increase. The range of point biserials (.22 to .62) is also reduced (see Table 1).

Insert Table 2 About Here

As deleting the misfitting items increases test reliability, change should manifest in the factor structure of the test items. Ryan and Hamm (1976) found that in such situations, the first factor became more pronounced.

> In particular, the proportion of total test variation accounted for by the first factor should increase after the items which do not fit the Rasch model has been deleted, because the remaining items will contain more true score information and will thus more accurately reflect the latent trait measured by the test (Ryan and Hamm, 1976, p. 2).

The first factor of the LMP-HMP instrument is pronounced (see Table 3). It behaves in a very different fashion from the remaining factors. For example, it is the only factor on which none of the items have negative loadings. In addition, the items which load on the factor have either negative or approximately zero loadings on the remaining factors. Thus, the factor structure of these items does lend supportive evidence to the concept of unidimensionality. Further, the pattern of loadings on factor 2 support the notion of a continuum of achievement. The LMP items tend to have negative loadings whereas the HMP tends to have positive loadings.

Insert Table 3 About Here

The factor structure of the revised LMP-HMP (see Table 4) test supports the observation of increased item homogeneity. And, the

Insert Table 4 About Here

loadings on the second factor reaffirm the unidimensionality of the test with LMP items tending to have negative loadings and the HMP items having (with one exception) positive loadings.

The Rasch analysis principal components factor, and the classical analysis each confirm the unidimensionality of this LMP-HMP instrument. Thus, it is more appropriate to describe this test as a unidimensional achievement continuum. One consequence of this empirical evidence attesting to the unidimensionality of Levin's instrument is the need for reanalysis of her study data. However, a more broad-reaching consequence of this analysis is the utility of psychometric analysis in the operationalization of multidimensional traits. The benefits of rigorous examination of item responses over conventional estimates of content validity are clear in this application. It is not sufficient to rely solely on content validity estimates and on response patterns to justify complex treatment of achievement measures. The results of this study suggest that when nonstandardized measures are the criteria for achievement, relying solely on estimates of content validity may lead to erroneous interpretation of the data.

References

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Kratwohl, D. R. (Eds.). *Taxonomy of educational objectives, Handbook I: Cognitive Domain*. New York: David McKay Co., Inc., 1956.

Fruckter, B. *Introduction to factor analysis*. New York: D. Van Nostrand Company, 1954.

Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.

Levin, T. *The effect of content pre-requisite and process-oriented experiences on application ability in the learning probability*. Unpublished Ph.D. dissertation, University of Chicago, 1975.

Lord, F. M. and Novick, M. R. *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley Publishing Co., 1968.

Ryan, J. P. and Hamm, D. W. *Practical procedures for increasing the reliability of classroom tests by using the Rasch model*. Paper read at the Annual Meeting of the National Council on Measurement in Education, San Francisco, Calif., 1976.

Weber, M. B. *The effects of environmental congruence on overt on-task behavior, on lower mental processes, and on higher mental processes*. Unpublished Ph.D. dissertation, University of Chicago, 1976.

Wright, B. and Mead, R. *Calfit: Sample-free item calibration with a Rasch measurement model*. Research memorandum No. 18. Chicago: Department of Education, University of Chicago, 1975.

Table 1

Rasch Item Analysis for the Original LMP & HMP Test

Item Name	Item Difficulty	Standard Error	Fit Mean Square	Original Point Biserial	Revised Point Biserial
1	-.701	.198	1.10	.34	.39
2	-.390	.196	3.20	.32	--
3	-1.590	.221	.65	.23	.26
4	-1.006	.203	2.05	.25	.22
5	1.577	.253	.68	.45	.45
6	-2.582	.286	.46	.32	.44
7	-0.475	.196	.95	.53	.57
8	-1.848	.233	.31	.25	.37
9	-1.062	.204	.50	.41	.46
10	-1.305	.211	.50	.36	.36
11	-1.671	.225	.50	.37	.48
12	2.164	.304	5.87	.02	--
13	-.308	.196	1.21	.58	.62
14	-.227	.196	1.14	.52	.57
15	-1.179	.207	.93	.38	.47
16	.588	.207	1.54	.60	.58
17	1.624	.256	1.10	.50	.48
18	-1.179	.207	1.07	.32	.36
19	.226	.200	2.85	.36	--

Table 1 - Cont'd.

Item Name	Item Difficulty	Standard Error	Fit Mean Square	Original Point Biserial	Revised Point Biserial
20	-.072	.197	1.85	.33	.39
21	.844	.215	1.61	.48	.51
22	.624	.208	1.57	.54	.54
23	-.149	.196	1.32	.54	.54
24	-1.945	.239	.66	.24	.33
25	.190	.199	2.53	.66	--
26	-.227	.196	1.17	.35	.36
27	.116	.196	3.50	.15	--
28	-.608	.197	.16	.43	.41
29	-.654	.198	.61	.46	.50
30	.444	.204	2.50	.30	--
31	--	--	0	.13	--
32	.770	.213	2.33	.28	--
33	--	--	0	.41	--
34	1.314	.237	2.12	.33	.34
35	--	--	0	.42	--
36	-.519	.197	.42	.42	.40
37	.733	.212	1.70	.43	.39
38	--	--	0	-.07	--
39	-.899	.201	.30	.31	.38
40	--	--	0	.08	--

Table 1 - Cont'd.

Item Name	Item Difficulty	Standard Error	Fit Mean Square	Original Point Biserial	Revised Point Biserial
41	--	--	0	.08	--
42	-1.119	.206	.63	.41	.46
43	-1.671	.225	.83	.40	.50
44	-1.371	.213	1.04	.48	.57
45	--	--	0	.27	--
46	--	--	0	.37	--
47	--	--	0	.26	--
48	1.231	.232	1.46	.28	.25
49	--	--	0	.01	--
50	.995	.221	0	.48	--
51	--	--	0	.36	--
52	--	--	0	.29	--
53	--	--	0	-.06	--
54	-1.757	.229	.69	.24	.35
55	.697	.211	1.20	.51	.51
56	1.929	.281	.76	.43	.41
57	-.563	.197	1.35	.37	.45
58	--	--	1.84	.29	.33
59	.995	.221	0	-.05	--
60	.004	.197	1.24	.41	.49
61	--	--	0	.20	--
62	--	--	0	.10	--
63	2.164	.304	1.71	.49	.47

Table 2

Summary Test Statistics

Test Name	Mean	Standard Deviation	Variance	KR_{20}
Original LMP-HMP (47 items)	29.31	8.68	75.29	.80
Revised LMP-HMP (40 items)	25.47	7.45	55.44	.89

Table 3

Factor Structure for Original LMP-HMP Test

Item Name	Estimates of Commonality	Factor 1 Loadings	Factor 2 Loadings
1	.49919	.37732	-.20399
2	.52210	.29539	-.27034
3	.60444	.23748	-.11183
4	.47858	.18041	-.01093
5	.68184	.44678	.20004
6	.50738	.43676	-.10131
7	.72151	.56490	-.28882
8	.52553	.36112	-.16182
9	.52364	.45813	-.25590
10	.63421	.38247	-.39385
11	.80261	.48299	-.54669
12	.46927	.01581	-.25394
13	.74196	.61762	-.41562
14	.67497	.57834	-.24473
15	.78181	.46268	-.54475
16	.65199	.60422	.01680
17	.67441	.49893	.09016
18	.50177	.38692	-.14438
19	.59473	.32758	.38514
20	.66867	.37458	-.17236
21	.47857	.48137	-.10648

Table 3 - Cont'd.

Item Name	Estimates of Commonality	Factor 1 Loadings	Factor 2 Loadings
22	.47072	.53499	-.00915
23	.62602	.56146	-.04487
24	.54045	.31383	.00726
25	.74204	.55103	.05591
26	.58899	.36508	-.23866
27	.40749	.14283	.00790
28	.49478	.40495	.01092
29	.62756	.47720	.19077
30	.64208	.28903	.36821
32	.53859	.28046	.38082
34	.47275	.30323	-.01527
36	.55751	.38189	.30036
37	.73594	.58777	.40972
39	.53148	.26991	.35568
42	.56948	.75396	.45578
43	.64879	.22007	.47780
44	.67719	.18486	.55271
48	.56981	.23603	.24415
50	.73846	.55106	.44098
54	.53911	-.04776	.32609
55	.58136	.11820	.48879

Table 3 - Cont'd.

Item Name	Estimates of Commonality	Factor 1 Loadings	Factor 2 Loadings
56	.52291	.15398	.40145
57	.59128	.03628	.39868
58	.55783	.09426	.27971
59	.62282	.02403	.43771
60	.56147	.16650	.47737
63	.42834	.00417	.33627

Table 4

Factor Structure for Revised LMP-HMP Test

Item Name	Estimates of Commonality	Factor 1 Loadings	Factor 2 Loadings
1	.39273	.36987	-.22383
3	.54663	.24367	.05936
4	.36527	.18938	.08110
5	.59080	.43107	-.02088
6	.46896	.44642	-.02305
7	.65766	.58380	-.34192
8	.50165	.37408	-.04131
9	.48145	.48384	-.18309
10	.55628	.40350	-.43258
11	.73997	.53127	-.45032
13	.72535	.64668	-.40641
14	.64238	.59289	-.23315
15	.73789	.51326	-.49606
16	.50462	.57913	-.06806
17	.61346	.48870	-.13496
18	.43714	.39100	-.07300
20	.62610	.39037	-.19009
21	.42441	.50352	-.11011
22	.42844	.53134	.01211
23	.49347	.53923	-.05699
24	.43289	.30636	.07436

Table 4 - Cont'd.

Item Name	Estimates of Commonality	Factor 1 Loadings	Factor 2 Loadings
26	.48821	.36175	-.02894
28	.42504	.38084	.04120
29	.60629	.47418	.40453
34	.40894	.29886	.04894
36	.39343	.36285	.35467
37	.51517	.31853	.22230
39	.47726	.34589	.32421
42	.53616	.44830	.45250
43	.60149	.48230	.47644
44	.62949	.55813	.37936
48	.47043	.21494	.04148
50	.51270	.33385	.10512
54	.48936	.47821	.06249
55	.44656	.37858	-.00105
56	.52038	.41261	.36671
57	.53625	.27355	.31040
59	.57490	.46018	.30309
60	.53250	.47142	.02780
63	.36039	.34429	.23377